

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2004, 13:13:40 ; Search time 3339 Seconds
(without alignments)
6788.981 Million cell updates/sec

Title: US-09-669-817A-4
Perfect score: 523
Sequence: 1 atctgacagcagatcacca.....gaccgcaagtggagtgga 523

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167151695 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rsd:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	515	98.5	1819	8 AK103829	AK103829 Oryza sat
2	515	98.5	194509	8 AP001129	AP001129 Oryza sat
3	467.6	89.4	1464	8 AK060638	AK060638 Oryza sat
4	130.2	24.9	1063	8 AK120505	AK120505 Oryza sat
5	130.2	24.9	1548	8 AK063929	AK063929 Oryza sat
6	130.2	24.9	1565	8 AK102953	AK102953 Oryza sat
7	130.2	24.9	110989	8 AP004049	AP004049 Oryza sat
8	130.2	24.9	135189	2 AP004851	AP004851 Oryza sat
9	109.2	20.9	7218	6 I66494	I66494 Sequence 14
10	108.2	20.7	63082	2 AC022663	AC022663 Homo sapi
11	107	20.5	303091	2 AC084799	AC084799 Mus muscu
12	106.6	20.4	265537	2 AC087228	AC087228 Mus muscu
13	106	20.3	69722	2 AC100560	AC100560 Mus muscu
14	106	20.3	252689	2 AC079433	AC079433 Mus muscu
15	105.8	20.2	3281	2 AC143602	AC143602 Macaca mu
16	105.4	20.2	203664	2 AC139759	AC139759 Mus muscu
17	105.2	20.1	81417	2 AC023265	AC023265 Homo sapi
18	105.2	20.1	81417	2 AC023265	AC023265 Homo sapi
19	105	20.1	222895	2 AC084065	AC084065 Mus muscu
20	104.8	20.0	60885	2 AC100962	AC100962 Mus muscu
21	104.8	20.0	164751	2 BX293567	BX293567 Danio rer
22	104.8	20.0	176734	2 AP003682	AP003682 Oryza sat
23	104.6	20.0	65505	2 AC027799	AC027799 Homo sapi
24	104.6	20.0	300695	2 AC079431	AC079431 Mus muscu
25	104.4	20.0	61223	2 AC130347	AC130347 Homo sapi
26	104.4	20.0	78220	2 AC023212	AC023212 Homo sapi
27	104.4	20.0	80291	2 AC101154	AC101154 Mus muscu
28	104.2	19.9	262980	2 AC098981	AC098981 Rattus no
29	104	19.9	126543	2 AC143806	AC143806 Macaca mu
30	104	19.9	261760	2 AC134459	AC134459 Mus muscu
31	103.8	19.8	239327	2 AL974309	AL974309 Danio rer
32	103.6	19.8	179206	9 CNS01DS6	AL121656 BAC seque
33	103.6	19.8	202083	2 AC023833	AC023833 Mus muscu
34	103.6	19.8	222707	2 AC079425	AC079425 Mus muscu
35	103.4	19.8	72847	2 AC023563	AC023563 Homo sapi
36	103.4	19.8	87283	2 AC022856	AC022856 Homo sapi
37	103.4	19.8	110737	2 AC011105	AC011105 Homo sapi
38	103.2	19.7	175146	2 AC132466	AC132466 Mus muscu
39	103	19.7	58483	2 AC100352	AC100352 Mus muscu
40	103	19.7	74138	2 AC021272	AC021272 Homo sapi
41	102.8	19.7	80133	2 AC021347	AC021347 Homo sapi
42	102.8	19.7	89994	2 AC021263	AC021263 Homo sapi
43	102.8	19.7	134940	2 AC018939	AC018939 Homo sapi
44	102.8	19.7	143733	2 AC142786	AC142786 Macaca mu
45	102.4	19.6	660	8 CNS019UW	AL112128 Botrytis

ALIGNMENTS

RESULT 1
AK103829
LOCUS
DEFINITION Oryza sativa (japonica cultivar-group) cDNA clone:J033148A06, full
insert sequence.
ACCESSION AK103829
VERSION AK103829.1 GI:32989038
KEYWORDS FLI_CDNA; CAP trapper.
SOURCE Oryza sativa (japonica cultivar-group)
ORGANISM Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.
REFERENCE
AUTHORS 1 The Rice Full-Length cDNA Consortium, National Institute of

Agrobiological Sciences Rice Full-Length cDNA Project Team:
Kikuchi,S., Satoh,K., Nagata,T., Kawagashira,N., Doi,K.,
Kishimoto,N., Yazaki,J., Ishikawa,M., Yamada,H., Ooka,H., Hotta,I.,
Kojima,K., Namiki,T., Ohneda,E., Yahagi,W., Suzuki,K., Li,C.,
Ohtsuki,K., Shishiki,T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group: Ootomo,Y., Murakami,K.,
Iida,Y., Sugano,S., Fujimura,T., Suzuki,Y., Tsunoda,Y.,
Kurosaki,T., Kodama,T., Masuda,H., Kobayashi,M., Xie,Q., Lu,M.,
Narikawa,R., Sugiyama,A., Mizuno,K., Yokomizo,S., Nishikura,J.,
Ikeda,R., Ishibiki,J., Kawamata,M., Yoshimura,A., Miura,J.,
Kusumegi,T., Oka,M., Ryu,R., Ueda,M., Matsubara,K., RIKEN:
Kawai,J., Carninci,P., Adachi,J., Alzawa,K., Arakawa,T., Fukuda,S.,
Hara,A., Hashidume,W., Hayatsu,N., Imotani,K., Ishii,Y., Itoh,M.,
Kagawa,I., Kondo,S., Konno,H., Miyazaki,A., Osato,N., Ota,Y.,
Saito,R., Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Yoshino,M., and Hayashizaki,Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice
JOURNAL
MEDLINE
PUBMED
22752273
12869764
2 (bases 1 to 1819)
REFERENCE
AUTHORS

Adachi,J., Alzawa,K., Akimura,T., Arakawa,T., Carninci,P., Doi,K.,
Fujimura,T., Fukuda,S., Hayashizaki,Y., Hayatsu,N., Hiramoto,K., Hiraoka,T.,
Hayashida,K., Hayashizaki,Y., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
Hori,F., Hotta,I., Iida,J., Iida,Y., Ikeda,R., Imamura,K.,
Imotani,K., Ishibiki,J., Ishii,Y., Ishikawa,M., Itoh,M., Kagawa,I.,
Kamagawa,S., Katoh,H., Kawagashira,N., Kawai,J., Kawamata,M.,
Kikuchi,S., Kishikawa-Hirozane,T., Kishimoto,N., Kobayashi,M.,
Kodama,T., Kojima,K., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Kurosaki,T., Kusumegi,T., Li,C., Lu,M.,
Masuda,H., Matsubara,K., Matsuyama,T., Miura,J., Miyazaki,A.,
Mizuno,K., Murakami,K., Murata,M., Nagata,T., Nakamura,M.,
Namiki,T., Narikawa,R., Nishikura,J., Nishi,K., Nomura,K.,
Numasaki,R., Ohneda,E., Ohno,M., Ohtsuki,K., Oka,M., Ooka,H.,
Osato,N., Ota,Y., Ootomo,Y., Ryu,R., Saitoh,H., Sakai,C., Sakai,K.,
Sakazume,N., Sano,H., Sasaki,D., Sato,K., Satoh,K., Shibata,K.,
Shinagawa,A., Shiraki,T., Shishiki,T., Sogabe,Y., Sugano,S.,
Sugiyama,A., Suzuki,K., Suzuki,Y., Tagami,M., Tagami-Takeda,Y.,
Tagawa,A., Takahashi,F., Takaku-Akahira,S., Tanaka,T., Tomaru,A.,
Toya,T., Tsunoda,Y., Ueda,M., Waki,K., Xie,Q., Yahagi,W.,
Yamada,H., Yamamoto,M., Yasunishi,A., Yazaki,J., Yokomizo,S. and
Yoshimura,A.
Direct Submission
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
Agrobiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,
Tel:81-29-838-7007, Fax:81-29-838-7007)
This clone is one of the 28K full-length cDNA clones from japonica
rice.
URL: http://cdna01.dna.affrc.go.jp/cDNA/
NIAS Rice Full-length cDNA Project Team: Kikuchi,S., Satoh,K.,
Nagata,T., Kawagashira,N., Doi,K., Kishimoto,N., Yazaki,J.,
Ishikawa,M., Yamada,H., Ooka,H., Hotta,I., Kojima,K., Namiki,T.,
Ohneda,E., Yahagi,W., Suzuki,K., Li,C., Ohtsuki,K., Shishiki,T. and
Yamamoto,M.
FAIS Genome Sequencing & Analysis Group: Ootomo,Y., Iida,Y.,
Fujimura,T., Ikeda,R., Ishibiki,J., Kawamata,M., Kobayashi,M.,
Kodama,T., Kurosaki,T., Kusumegi,T., Lu,M., Masuda,H., Miura,J.,
Mizuno,K., Narikawa,R., Nishikura,J., Oka,M., Ryu,R., Sugano,S.,
Sugiyama,A., Suzuki,Y., Tsunoda,Y., Ueda,M., Xie,Q., Yokomizo,S.,
Yoshimura,A., Matsubara,K. and Murakami,K.
Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken: Adachi,J., Alzawa,K.,
Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hanagaki,T.,
Hara,A., Hashizume,W., Hayashida,K., Hayatsu,N., Hiramoto,K.,
Hiraoka,T., Hori,F., Iida,J., Imamura,K., Imotani,K., Ishii,Y.,
Itoh,M., Kagawa,I., Kanagawa,S., Katoh,H., Kawai,J.,
Kishikawa-Hirozane,T., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Ohno,M., Osato,N.,
Ota,Y., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H.,

COMMENT
TITLE
JOURNAL

Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Tagami,M., Tagami-Takeda,Y., Tagawa,A., Takahashi,F.,
Takaku-Akahira,S., Tanaka,T., Tomaru,A., Toya,T., Waki,K.,
Yasunishi,A. and Hayashizaki,Y.
Location/Qualifiers
1..1819
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J033148A06"
ORIGIN
Query Match 98.5%; Score 515; DB 8; Length 1819;
Best Local Similarity 99.0%; Pred. No. 1.8e-66;
Matches 518; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 ATCTGATCAGCAGATCACCAGTCTCTTCACTCCACTCCAGCGGAGCTGAGAGA 60
DB 17 ATCAGATCAGCAGATCACCAGTCTCTTCACTCCACTCCAGCGGAGCTGAGAGA 76
QY 61 GAGCTCCACCGCGCGGCCACCATGAGTCTCTACACCGCTTACGCCACCTCTGAGC 120
DB 77 GAGCTCCACCGCGCGGCCACCATGAGTCTCTACACCGCTTACGCCACCTCTGAGC 136
QY 121 ACCCTTCTTCCCTTCCACCGCGCGCTCTCTCTCTGCCCCCTTCTCTCCACCCG 180
DB 137 ACCCTTCTTCCCTTCCACCGCGCGCTCTCTCTCTGCCCCCTTCTCTCCACCCG 196
QY 181 CCGTTTCTTCCACCTGCGCATTTCTGCGCTCGACTCCCCCTTCTGCTGAGCCCTTCC 240
DB 197 CCGTTTCTTCCACCTGCGCATTTCTGCGCTCGACTCCCCCTTCTGCTGAGCCCTTCC 256
QY 241 ATCTCCACCCCTTCTCTCCCAACCGCGGACCACTAGTCTCTCTCTGAGCCCTTCTCTCC 300
DB 257 ATCTCCACCCCTTCTCTCTCCCAACCGCGGACCACTAGTCTCTCTCTGAGCCCTTCTCTCC 316
QY 301 ACACCTTCCACCGACCGCGCTCTCCCACTGAGCTGCGCTCTCTCTGAGCCCTTCTCTCC 360
DB 317 ACACCTTCCACCGACCGCGCTCTCCCACTGAGCTGCGCTCTCTCTGAGCCCTTCTCTCC 376
QY 361 CCAGGCCCCACGACGAGAGTGCACCTATGTTCACCGAGTCCACCGCGCGCATGTCAAGT 420
DB 377 CCAGGCCCCACGACGAGAGTGCACCTATGTTCACCGAGTCCACCGCGCGCATGTCAAGT 436
QY 421 GGACACCGGAGACAAAGCTTGGCGCGCGGAGAGGGTCTCAAGTGGAGGCTGAGCTCG 480
DB 437 GGACACCGGAGACAAAGCTTGGCGCGCGGAGAGGGTCTCAAGTGGAGGCTGAGCTCG 496
QY 481 ACTCCCTTACGACGAGTGAATTCGACCGCAAGTGAAGTGGGA 523
DB 497 ACTCCCTTACGACGAGTGAATTCGACCGCAAGTGAAGTGGGA 539
RESULT 2
AP001129/c 194509 bp DNA linear PLN 21-MAR-2002
DEFINITION Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 6,
PAC clone:P0644B06.
ACCESSION AP001129
VERSION AP001129.1 GI:6907081
KEYWORDS
SOURCE
ORGANISM
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
1 Sasaki,T., Matsumoto,T. and Yamamoto,K.
Oryza sativa nipponbare(GAS) genomic DNA, chromosome 6, PAC
clone:P0644B06
2 (bases 1 to 194509)

AUTHORS Sasaki, T., Matsumoto, T. and Yamamoto, K.
 TITLE Direct Submission
 JOURNAL Submitted (02-FEB-2000) Takuji Sasaki, National Institute of
 Agrobiological Resources, Rice Genome Research Program, Kannondai
 2-1-2, Tsukuba, Ibaraki 305-8602, Japan
 (E-mail:tsasaki@abr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
 Tel:81-298-38-7441, Fax:81-298-38-7468)
 COMMENT The orientation of the sequence is from T7 to SP6 of the PAC clone.
 Genes were predicted from the integrated results of the
 following: GENSCAN1.0, BLASTN2.0, BLASTX2.0 as well as
 SplicePredictor (October 1998 version). The genomic sequence was
 searched against the non-redundant database NRP (PIR, SWISSPROT,
 GENPEPT, PDB) from MAFB DNA bank and the cDNA sequence database at
 RGP. Protein similarities of the coding regions were searched
 against NRP with BLASTP2.0. ESTs represent the identified cDNA
 sequences using BLASTN 2.0 with the corresponding DDBJ accession
 no. and RGP clone ID.
 This sequence of this clone has an overlap with P0514G12 clone,
 DDBJ:AP000616 at the 3' end. This clone ends at the position 30,063
 of P0514G12. Detailed information on overlap and assembly quality
 together with annotation of this entry at
 http://www.dna.affrc.go.jp:82/genomicdata/genomefinished.html.
 location/Qualifiers
 1. 194509
 /organism="Oryza sativa (japonica cultivar-group)"
 /mol_type="genomic DNA"
 /cultivar="Nipponbare"
 /db_xref="taxon:39947"
 /chromosome="6"
 /clone="P0644B06"
 join(53. 107,495. 542,2129. 2223)
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA90609.1"
 /db_xref="GI:6907082"
 /translation="MACOYCYKLRILTVMYVPHITVSEKVLIMAAACSSYLARYLQV
 QIPRRMLQKHVQGHETSSAKV"
 complement(join(6634. 6742,6856. 7058,7301. 7379,
 7465. 7830,8543. 8909,9188. 9326,9414. 9684,9802. 9911,
 10520. 10657))
 /note="ESTs AU082304(C61278), C93463(C50982) correspond to
 a region of the predicted gene.
 Similar to Arabidopsis thaliana DNA chromosome 4, BAC
 clone F24A6, Pollen-specific protein precursor like.
 (AL035396)"
 /codon_start=1
 /protein_id="BAA90610.1"
 /db_xref="GI:6907083"
 /translation="MAAALAFVAVAGALVAOATDPYVFPDWDVSFITASPLGLP
 QXVIALNKQPPGPVNLITNNVNVVNLNSLDEPLLITWDGIQKNCWQDGLGTC
 PIPGMMTYNFQVKDQIGSEFYEPPLSLQRAAGFGITVNNRAVISVPDTPDGI
 TLFIDGWYKSHSTDLRKLDDGKELGMPDGYLMNGKGPYRYNDSLVPAGIBHETIKVE
 PGKYRFRVHNVGISTSLNRIQNNHIALVEEGSYTMKQNFNDIHVGOSYSLIT
 MDONASSDYIVASARFVNESLMTKVTAIIQYNSKSKAGSLDPDPNDEYDKTFS
 MNQARSIRMNVTGAARBNPQGSFHYGINSQVYKLRNEPPTVINGKRTTILSGISF
 SPDPDPMRLADLYDKKGYTLDPPTMPIDGPVMKTSVINSTYKNFLIVFONNDTKV
 QYHIDGYAFWVVGMDYGEWTEENSRGTYNKMGVSRCTQVFPQAWTAVMLSLDSPGF
 WNVRTENLDTWYLQGETYIRIVDPPDGGYNVTEWVAPDNMLYGLLKDQKAKPKHSS
 SSSSSAASPNRYWLAVVSLVAAVFVQ"
 complement(join(11097. 11247,12146. 12162))
 /note="hypothetical protein"
 /codon_start=1
 /protein_id="BAA90611.1"
 /db_xref="GI:6907084"
 /translation="MSHVNSTGARHGTAIVAAPGTARGTARPVSPARAQESATWAG
 GGRMGRRHVAH"
 complement(join(15234. 15297,15414. 15616,16201. 16476,
 16777. 17562))
 /note="ESTs AU082316(E3368), D41461(S3973) correspond to a
 region of the predicted gene.
 hypothetical protein"
 /codon_start=1
 /protein_id="BAA90612.1"

```

CDS
    /db_xref="GI:6907085"
    /translation="MSSHRYAHLDDPFFPPPPSSSCPFLSPPAASSTCFPALD
    /SPFADPFLHPEDLPPTSSLDPLFHLTLTRVSGLELAARAAPHRTSRKCTY
    VTESTGRKVKMTTEDKPRAGERVLKWEAEIDSPYDSDGRKMKWEAKATASAAATKL
    KWATHLKGKCLEPWSHSYTWEEDFSATDDDDDEETEDQLHKALQDHSKLTKAKOD
    KKKKKDNNTVVVNKEQKKCPFSVKIEIIPPEEDNTAGCAIRKAFALNGKAKKEL
    SPQDAALLIOLNRYRAHLAHSQVLRCLDLAVAKAKLEIRLSLFYNISYRHMAHDE
    ERQFTEKILVLLTLVDALBEGPDYVRTAKKSMLELEGMLEIVDPQPPGKQSLTR
    KFDLPFGGPITDEKMGAVNVKVIQKXNLRLNTLCTWHGSLIVLASKH"
    join(20129, 20894, 20952, 21196, 21423, 21833)
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAA90613.1"
    /db_xref="GI:6907086"
    /translation="MGRPAOKAESAARPKGERETVDRAHLAWSRVGAPCQRLGSPCG
    GSMRRALFGEGETRCTRAVGRCGGPTSLTARGCPRLRAGEGAQDGRDLDPVLA
    ELAPTRLRGCGHAGREVDGAGRNGRTTAAGCANHGDTGSEHGTMLHETRGEEB
    TARIRRELDDGELRRQPAEEGGNGDEVTRGRFPAPASTRLRESVASVGLGAVP
    SEAGDERFRARAATASTRAMAAGAPASYGAEATRTTVRAGAVTRADSAKDRFC
    LGPRKRRKRAPSPSASLLVPAPTRATNGDEOGSGGERRKGRFGGGAWVYR
    GGNRVEMQGPDDDDGDDGGLSGAARARAVLGAERLRGGRGRAGAAGPDAEGAA
    RAATTRARAEORRAGAEGESGARARARARAGARAERVRSPGRRERESRGGRGWAE
    RDSAHPTGRNRLLRDRFREGL"
    22425, 23416
    /note="5' LTR"
    join(24954, 25336, 25545, 26190, 26729, 26833)
    /note="ESTs C72771(E2215), AU082313(E2215) correspond to a
    region of the predicted gene.
    hypothetical protein"
    /codon_start=1
    /protein_id="BAA90614.1"
    /db_xref="GI:6907087"
    /translation="MAPPIAGNGRAKSGRLPLPPAIPVAVPPSPVLGSLPSAIKCNP
    LAPFRFSPLRSVPVPSLPVALLSLPCRPPPAVSPSPERRPAPSPPASQLP
    LPRLALRFPRRPELRSRRRPSLSARCRPLPRHGVALHLVHPSVSPGSRVATNP
    STRLAGVLRPPVPRFLTQOPRHHLRHSKQVALGLASPPSKPLRGSSSSPILVV
    DVPSVTRLVSSRRRLRRRHSRGVHAVGCFHRLLAGPGRRRRLVPPGRGVPSV
    ELFSSVLGVSSRLQPPSSSSSPRRQASCPRLVAVGSPPKSPFRLSLPRLFVV
    VVPTPRRVVACSFACVLVAFVPEVPEAWFVVAEGSGRSL"
    join(28657, 29959, 30038, 32303, 33392, 34484)
    /note="Similar to Sorghum bicolor 22 kDa kafirin cluster,
    polyprotein. (AF061282)
    internal stop codon:(31792, 31794)
    probably inactive because stop codons are included in CDS"
    /pseudo
    /codon_start=1
    34828, 35824
    /note="3' LTR"
    complement(join(36178, 36567, 37061, 37297, 37473, 37610))
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAA90615.1"
    /db_xref="GI:6907088"
    /translation="WTLILTRIQPVGISHVGRPPSFHDKAFRKPOTVYHMPTRGVDPQ
    QOETVLKCKSGKPYTRYVPRIASVLEIHLVRKHPRCFVQVCFCKPKFYPSCYDCKFE
    GDRRYSHRVREYRGAAGWLQDGMGAQEKWSGPAELNGPAPRRPKLARAREGEREP
    VDRAHHARSRVGAPCQRLGSPCGGSTRTRAGGEGTRCMRAVRRGRCGGPTSLTAR
    GGSRRAGGTDRGRPDPVWLWRRRGAYVAATRAGGR"
    38322, 39598
    /note="5' LTR"
    join(40025, 41255, 41586, 41833, 41855, 42049, 42113, 42486
    42670, 42804, 42980, 43171, 43364, 43895, 44076, 445490,
    45572, 45842)
    /note="Similar to Sorghum bicolor Gypsy-Ty3 type
    retrotransposon RetroSor1, polyprotein. (AF098806)
    internal stop codon:(40841, 40843)
    probably inactive because stop codons are included in CDS"
    /pseudo
    /codon_start=1
    complement(join(46472, 48407, 48502, 49763))
    /note="hypothetical protein"
    /codon_start=1
    /protein_id="BAA90616.1"

```

```

/db_xref="GI:6907089"
/translation="MAGSALLDGLVLPSPRIVSERQAGIPRRFMESATGREIVTLG
EGRPAPCPGRSVFELTFPAMAGLVPESSFFMDVLEFYDQMAHLTPNAVMTLAFAH
LCMEFIVRPSLRLFRWFFTVSPSPVGGCYFQPRGVLNRYIPCVLRRKMDWK
SDMFYTPLADEARLRLPSQPAQSSWRAVVDIGDYADLRLAGLRSGLTGMVFA
GDYLRRIAPLQRRARGAMEYTGSEDMYRTHQGRWDWAPEDFKIVQRLNLSSVEA
SLIPQVPLCSDPDRASTLITMQAVGASERAPRGHDGAGSRGEQSTPGGGRASG
PRNGGPGGRPADARQKQKQEPSPRGVAPRASHSGQSEADPATTEARRRE
RPHXMGTEPSRGNLISPPRWSFNPRSEI PSRPSRHSQSGSEADPATTEARRRE
SDRREADRLREAEAAQAVRARQAEAREARLQAEATTTSETADEAAGASLGTPSCAR
RLREAEAAQAVRARQTEEAAREARLQAEATTTSETADEAAGASLGTPSCAR
DKPGPDIPESGTSIGSPRAASSRLPTPSVAPLSAEPPLQALAVANTMVLGDS
AQMEALQAEVVELDAMARVEEGRSVEAMVEVGRKARHRSVSELEARKVLGETAKE
VEERGAALIAATVMNEADOTLRLOYGWEETELGKLDARGLDAAARERAAETE
VASRRREALEARAMALEERACAVEADLADREAAVAIREATLTAHEACAEEESALRL
REDALTERERALEAEAAQRLADSLREAVREEOARNEEDARAERALNORAEEL
EAREKELDARARSGGAAGESDITARLAAEHITADLQALDSSAGEVEALRLAGEVG
PGMLRDVSRDRAGROAGLWGRGRTAKYANQRLAQLSEIAGTLQRLPELEERTIK
SSSRDLARGAVELVLASYQARDPDPSPWTALDEFPPTGTDGARAQVRDAADHIVHSFE
GSAPRLAFALNSDEEDDDGVDSDGEADDPGASE"
50029.51305
/contig="3", LTR"
join(53110..53450,54047..54608)
/note="Similar to Oryza sativa root-specific RCC3 mRNA.
(L27208)"
/codon_start=1
/protein_id="BAA90617.1"
/db_xref="GI:6907090"
/translation="MAGVACVLVGVLMISAGQSAACDGHPCPTPAGKCPINTVKLG
CADVLGLIHASTPPEKPCPLIAGLADLDAVCLAINANLGLNDVPLDLSLL
NYCGCKLPAGFNTGGGCMWRGAAAPPPDPAGGEARQVAQEGRRRAKGGCS
GWRISLPQVAAWRQRRRRPRLRQIRWEGRRRAPAGALPKXAAAAVSPPPDL
AGETASSRGGCLPGTAAAAAASPPLDPAEGEAVGRRRAPAGGLEEAAAAHSGG
GRAIADLKKGKAGGNPREKDEGS"
complement(56226..57308)
/note="Similar to Rice mRNA for lactate dehydrogenase.
(D13817)"

```

CDS
/note="Similar to Rice mRNA for lactate dehydrogenase.
(D13817)"

```

Query Match      98.5%; Score 515; DB 8; Length 194509;
Best Local Similarity 99.0%; Pred. No. 6.2e-67;
Matches 518; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ATCTGATCAGAGATCACCAAGTCTCTTCCACTCCCACTCCAGCGGAGCTGAGCAGA 60
Db 17645 ATCAGATCAGAGATCACCAAGTCTCTTCCACTCCCACTCCAGCGGAGCTGAGCAGA 17586

QY 61 GAGCTCCACCGCGCGCGGACCATGAGTCTTACCAACCGTTACGCCACCTCTCGACG 120
Db 17585 GAGCTCCACCGCGCGCGGACCATGAGTCTTACCAACCGTTACGCCACCTCTCGACG 17526

QY 121 ACCCTTCTTCCCTTCCACCGCGCTCTCTCTCTCTGCCCCCTTCTCTCTCCACCCG 180
Db 17525 ACCCTTCTTCCCTTCCACCGCGCTCTCTCTCTCTGCCCCCTTCTCTCTCCACCCG 17466

QY 181 CCGTTCTCTCCACTGCCCCATTTCTGCCCCCTGACTCCCCCTTGGTGTGACCTTTCC 240
Db 17465 CCGTTCTCTCCACTGCCCCATTTCTGCCCCCTGACTCCCCCTTGGTGTGACCTTTCC 17406

QY 241 ATCTCCACCCCTTCTCTCCACCGCGGACCACTAGTCTCTCTCTGACCCCTTCTCTCC 300
Db 17405 ATCTCCACCCCTTCTCTCCACCGCGGACCACTAGTCTCTCTCTGACCCCTTCTCTCC 17346

QY 301 ACACCTTCCACCGACCGCGTCTCCCACTGAGTCTGCGCTGCGCGCGCGCGCCACACC 360
Db 17345 ACACCTTCCACCGACCGCGTCTCCCACTGAGTCTGCGCTGCGCGCGCGCGCCACACC 17286

QY 361 CCAGGCGCCACGACGAGAGTGCACCTATGTACCGAGTCCACCGCGCGCATGTGCAAGT 420
Db 17285 CCAGGCGCCACGACGAGAGTGCACCTATGTACCGAGTCCACCGCGCGCATGTGCAAGT 17226

QY 421 GGACCCACGAGACAGGCTCGCGCGCGCGGAGAGGAGTGTCTCAAGTGGAGGCTGAGCTCG 480
Db 17225 GGACCCACGAGACAGGCTCGCGCGCGCGGAGAGGAGTGTCTCAAGTGGAGGCTGAGCTCG 17166

```

```

QY 481 ACTCCCCCTACGACGATGATTCGACCCGCGAAGTGGAAGTGCGGA 523
Db 17165 ACTCCCCCTACGACGATGATTCGACCCGCGAAGTGGAAGTGCGGA 17123

RESULT 3
AK060638
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS

```

```

1
The Rice Full-Length cDNA Consortium, National Institute of
Agrobiological Sciences Rice Full-Length cDNA Project Team;
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
Ohtsuki, K., Shishiki, T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group; Ohtsuki, T., Murakami, K.,
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Mikura, J.,
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
Kusunegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN;
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,
Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Yoshino, M., and Hayashizaki, Y.
Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice
Science 301 (5631), 376-379 (2003)
22752273
12869764
2 (bases 1 to 1464)
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,
Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K.,
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Kurosaki, T., Kusunegi, T., Li, C., Lu, M.,
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,
Namiki, T., Narikawa, R., Mikura, J., Nishi, K., Nomura, K.,
Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H.,
Osato, N., Ota, Y., Ohtsuki, T., Ohtsuki, K., Ohtsuki, K., Sakai, K.,
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and
Yoshimura, A.
Direct Submission
Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of
Agrobiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression; 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp,
Tel:81-29-838-7007, Fax:81-29-838-7007)
This clone is one of the 28K full-length cDNA clones from japonica
rice.
URL : http://cdna01.dna.affrc.go.jp/cDNA/

```


NIAS Rice Full-Length cDNA Project Team: Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T. and Yamamoto, M.

FAIS Genome Sequencing & Analysis Group: Ootomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Nikura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

FEATURES

Location/Qualifiers
1. 1464
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="001-027-D03"

ORIGIN

```
Query Match      89.4%; Score 467.6; DB 8; Length 1464;
Best Local Similarity 99.2%; Pred. No. 1.7e-59;
Matches 470; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 50 AGCTGAGCAGAGAGCTCCACCGCGCGGCCACCATGAGTTCTCACCACCGTTACGCCCA 109
DB 1 AGCTGAGCAGAGAGCTCCACCGCGCGGCCACCATGAGTTCTCACCACCGTTACGCCCA 60

OY 110 CTTCTCGACGACCGCTTCTTCCCTTCCACCGCGCGTCTCTCTCTGCGCCCTTCT 169
DB 61 CTTCTCGACGACCGCTTCTTCCCTTCCACCGCGCGTCTCTCTCTGCGCCCTTCT 120

OY 170 CTCCCCACCGCGCGTTTCTCTCCACCGCGCGCATTTCTTGCCCTCGACTCCCTCGTGC 229
DB 121 CTCCCCACCGCGCGTTTCTCTCCACCGCGCGCATTTCTTGCCCTCGACTCCCTCGTGC 180

OY 230 TGACCTTTTCATCTCCACCGCTTCTCTCCACCGCGCGCATTAAGTCTCTCTCGACCC 289
DB 181 TGACCTTTTCATCTCCACCGCTTCTCTCCACCGCGCGCATTAAGTCTCTCTCGACCC 240

OY 290 CTTCTCTCTCCACCGCTTCTCCACCGCGCGTCTCTCCACTCGAGTGGCCCTCGCGCCG 349
DB 241 CTTCTCTCTCCACCGCTTCTCCACCGCGCGTCTCTCCACTCGAGTGGCCCTCGCGCCG 300

OY 350 CGCCCCACACCCGAGGCCACCGAGAGAGTGACCTATGTGACCGAGTCCACCGCGCG 409
DB 301 CGCCCCACACCCGAGGCCACCGAGAGAGTGACCTATGTGACCGAGTCCACCGCGCG 360

OY 410 CATGTCAGGTGAGCACCAGAGACCAAGCTCGCGCGCGGAGAGGGTGTCTCAAGTGGA 469
DB 361 CAAGTCAAGTGAGCACCAGAGACCAAGCTCGCGCGCGGAGAGGGTGTCTCAAGTGGA 420

OY 470 GGCTGAGCTGACTCCCTTACGAGAGATGATTCGACCGCAAGTGAAGTGGA 523
DB 421 GGCTGAGCTGACTCCCTTACGAGAGATGATTCGACCGCAAGTGAAGTGGA 474
```

RESULT 4
AK120505/c

LOCUS

AK120505 1063 bp mRNA linear PLN 29-OCT-2003
Oryza sativa (japonica cultivar-group) cDNA clone:J013123A20, full
insert sequence.
AK120505
AK120505.1 GI:37990128
FLI CDNA; CAP trapper.

KEYWORDS

Oryza sativa (japonica cultivar-group)

SOURCE

Oryza sativa (japonica cultivar-group)

ORGANISM

Oryza sativa (japonica cultivar-group)

REFERENCE

1

The Rice Full-Length cDNA Consortium, National Institute of

Agrobiological Sciences Rice Full-Length cDNA Project Team;

Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,

Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,

Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,

Ohtsuki, K., Shishiki, T., Foundation of Advancement of International

Science Genome Sequencing & Analysis Group; Ootomo, Y., Murakami, K.,

Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,

Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,

Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nikura, J.,

Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,

Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN;

Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,

Hara, A., Hashizume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,

Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,

Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,

Yoshino, M. and Hayashizaki, Y.

Collection, mapping, and annotation of over 28,000 cDNA clones from

Japonica rice

Science 301 (5631), 376-379 (2003)

22752273

12869764

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

JOURNAL

Medicine

Science

301 (5631), 376-379 (2003)

22752273

12869764

2

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,

Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,

Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,

Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K.,

Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,

Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,

Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,

Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,

Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,

Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,

Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, Y.,

Nakamura, M., Namiki, T., Narikawa, R., Nikura, J., Nishi, K.,

Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M.,

Ooka, H., Osato, N., Ota, Y., Ootomo, Y., Ryu, R., Saitoh, H., Sakai, C.,

Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K.,

Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y.,

Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M.,

Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S.,

Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K.,

Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J.,

Yokomizo, S. and Yoshimura, A.

Collection, mapping, and annotation of 28k full-length cDNA clones

from japonica rice

Unpublished

3 (bases 1 to 1063)

Kikuchi, S.

Direct Submission

Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of

Agrobiological Sciences, Department of Molecular Genetics, Head of

laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki

305-8602, Japan (E-mail:skikuchi@nias.affrc.go.jp,

Tel:81-29-838-7007, Fax:81-29-838-7007)

This clone is one of the 32k full-length cDNA clones from japonica

rice.

URL: http://cdna01.dna.affrc.go.jp/cdna/

NIAS Rice Full-Length cDNA Project Team: Kikuchi, S., Satoh, K.,

Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J.,

Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Namiki, T.,

Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Yamamoto, M. and Nakahama, Y.
FAIS Genome Sequencing & Analysis Group: Otsomo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Ninkura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S., Yoshimura, A., Matsubara, K. and Murakami, K. Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Osato, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takeku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

FEATURES

```
1..1063
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="J013123A20"
```

ORIGIN

Query Match	24.9%	Score 130.2;	DB 8;	Length 1063;
Best Local Similarity	58.9%	Pred. No. 5.4e-10;		
Matches 269; Conservative	0;	Mismatches 173;	Indels 15;	Gaps 2;

QY	79	CCACCATGAGTTCTCACCACCGGTAGGCCACCTCCTCGACGACCCCTTCTTCCCTTCC	138
Db	626	CCACCTTCCCAACCTCTCTCTCTCTCTCTCGACGACACCCACCATTTTCTTCC	567
QY	139	CACCGCGCTCTCTCTCTCTCTGCCCCCTCTCTCTCCCAACCGCGCTTCTCTCGACCTGCC	198
Db	566	CTCCCTCTCTCTCATGCCCCACTCGGTTCACTCTCTCTCTCATGCTTGCCCACTCG	507
QY	199	CATTCTTCCGCTCGACTCCCCCTTGCCTGTGACCCCTTTCATCTCCACCCCTTCTCC	258
Db	506	GCTTCACTCTCTCTACGACCTCGACGCGCCTTCCACCAACCATCTCGACCTCG	447
QY	259	CCACG-----CCGACCACTAGCTCCCTCTCGACCCCTTCTCTCTCCACACC	306
Db	446	ACCTTGACCTTCTCTCCCAACCCCGACCCACCTGTAAACACCTCTGCCCGCCCTCC	387
QY	307	TCAACGACCGCGTCTCCCACTCGAGCTCGCCCTCGCGCGCGGCCACACCCCAAGC	366
Db	386	GCGATCCCTTCTCTCTCCACCACTCCCTCGCCACCGCGCTCTCCGCTTGAGCTCGCGG	327
QY	367	CCACCAAGCAAGTGCACCTATGTCAACCGAGTCCACCGGCGGCAATGTCATAGTGCACCA	426
Db	326	CCCCACGCAAGTACACCTACGAGCGCGAATCCGCCGCGCAGGAAGATCAAGTGCACCA	267
QY	427	CCGAGGACAAGCCTCGCGCGCGCGAGAGGGTCTCAAGTGGAGGCTGAGCTCGACTCCC	486
Db	266	CCCAACGCAAGCC--CGACGCGGACCGGACCTTCAAGTGGGAGGCCCAAGATGCACACC	210
QY	487	CCTACGACGATGATTGCAACCGCAAGTGAAGTGGGA	523
Db	209	CCAAACGACGATGGCTTTCGACCGCGAAGTGAAGTGGGA	173

RESULT 5	AK063929	LOCUS	AK063929	1548 bp	mRNA	linear	PLN 24-JUL-2003
DEFINITION	Oryza sativa (japonica cultivar-group) cDNA clone:001-123-C12, full						

ACCESSION

insert sequence.

AK063929

VERSION

FLI_CDNA; oligo-capping.

SOURCE

Oryza sativa (japonica cultivar-group)

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzae; Oryza.

REFERENCE

The Rice Full-Length cDNA Consortium, National Institute of

AUTHORS

Agrobiological Sciences Rice Full-length cDNA Project Team:
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
Ohtsuki, K., Shishiki, T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group: Otomo, Y., Murakami, K.,
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Ninkura, J.,
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN;
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
Hara, A., Hashidume, W., Hayatsu, N., Imocani, K., Ishii, Y., Itoh, M.,
Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Yoshino, M. and Hayashizaki, Y.

TITLE

JOURNAL Science 301 (5631), 376-379 (2003)

MEDLINE

12869764

REFERENCE

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,

AUTHORS

Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imocani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, C., Kojima, T., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurossaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M., Namiki, T., Narikawa, R., Nikura, J., Nishi, K., Nomura, K., Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Osato, N., Ota, Y., Otomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaki-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.

TITLE

Submitted (05-DEC-2001) Shoshi Kikuchi, National Institute of

JOURNAL

Agrobiological Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: sktkuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)

COMMENT

This clone is one of the 28K full-length cDNA clones from japonica rice.

URL : http://cdna01.dna.affrc.go.jp/cdna/NAS_Rice_Full-length_cDNA_project/Team:kikuchi,S.,Sato,H.,Nagata,T.,Kawagashira,N.,Doi,K.,Kishimoto,N.,Yazaki,J.,Ishikawa,M.,Yamada,H.,Ooka,H.,Hotta,I.,Kojima,K.,Namiki,T.,Ohneda,E.,Yahagi,W.,Suzuki,K.,Li,C.,Ohtsuki,K.,Shishiki,T. and Yamamoto,M.

Fais Genome Sequencing & Analysis Group: Ocampo, Y., Iida, Y., Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M., Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J., Mizuno, K., Narikawa, R., Nukura, J., Oka, M., Ryu, R., Sugano, S., Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.

Yoshimura, A., Matsubara, K. and Murakami, K.
Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
Akimura, T., Arikawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
Hirooka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
Ito, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M.,
Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Oseto, N.,
Ota, Y., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H.,
Sasaki, D., Sato, K., Shibata, K., Shingawa, A., Shiraki, T.,
Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F.,
Takaku-Akhlira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K.,
Yasunishi, A. and Hayashizaki, Y.

```

FEATURES
    source
        location/Qualifiers
            1..1548
                /organism="Oryza sativa (japonica cultivar-group)"
                /mol_type="mRNA"
                /cultivar="Nipponbare"
                /db_xref="taxon:39947"
                /clone="001-123-C12"

```

ORIGIN

Query Match	24.9%;	Score 130.2;	DB 8;	Length 1548;
Best Local Similarity	58.9%;	Pred. No. 5e-10;		
Matches 269;	Conservative 0;	Mismatches 173;	Indels 15;	Gaps 2;

[illegible]

RESULT 6
AK102953
LOCUS
DEFINITION
AK102953
Oryza sativa (japonica cultivar-group) 1565 bp mRNA linear PLN 24-JUL-2003
insert sequence. cDNA clone:J033115E12, full
insert sequence.
AK102953
AK102953.1 GI:32988162
FLI_CDNA; CAP trapper.
Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

REFERENCE
AUTHORS

TITLE

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

1 Ehrhartoideae; Oryzeae; Oryza.

1 The Rice Full-Length cDNA Consortium, National Institute of
Agrobiological Sciences Rice Full-Length cDNA Project Team;
Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K.,
Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I.,
Kojima, K., Namiki, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C.,
Ohtsuki, K., Shishiki, T., Foundation of Advancement of International
Science Genome Sequencing & Analysis Group; Otomo, Y., Murakami, K.,
Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y.,
Kurosaki, R., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M.,
Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Ninkura, J.,
Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J.,
Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., RIKEN;
Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S.,
Hara, A., Hashidume, W., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M.,
Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Osato, N., Ota, Y.,
Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T.,
Yoshino, M., and Hayashizaki, Y.

Collection, mapping, and annotation of over 28,000 cDNA clones from
japonica rice
Science 301 (5631), 376-379 (2003)

22752273
12869764

2 (bases 1 to 1565)

Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K.,
Fujimura, T., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W.,
Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hiraoka, T.,
Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K.,
Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I.,
Kanagawa, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M.,
Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M.,
Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,
Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M.,
Masuda, H., Matsubara, K., Matsuyama, T., Miura, J., Miyazaki, A.,
Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakamura, M.,
Namiki, T., Narikawa, R., Ninkura, J., Nishi, K., Nomura, K.,
Numasaki, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H.,
Osato, N., Ota, Y., Otomo, Y., Ryu, R., Saitoh, H., Sakai, C., Sakai, K.,
Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K.,
Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S.,
Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami-Takeda, Y.,
Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A.,
Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W.,
Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and
Yoshimura, A.

Direct Submission
Submitted (27-AUG-2002) Shoshi Kikuchi, National Institute of
Agrobiological Sciences, Department of Molecular Genetics, Head of
Laboratory of Gene Expression, 2-1-2 Kannondai, Tsukuba, Ibaraki
305-8602, Japan (E-mail: skikuchi@nias.affrc.go.jp,
Tel:81-29-838-7007, Fax:81-29-838-7007)

This clone is one of the 28K full-length cDNA clones from japonica
rice.
URL: http://cdna01.dna.affrc.go.jp/cDNA/NIAS_Rice_Full-Length_cDNA_Project_Team_Kikuchi_S_Satoh_K_Nagata_T_Kawagashira_N_Doi_K_Kishimoto_N_Yazaki_J_Ishikawa_M_Yamada_H_Ooka_H_Hotta_I_Kojima_K_Namiki_T_Ohneda_E_Yahagi_W_Suzuki_K_Li_C_Ohtsuki_K_Shishiki_T_and_Yamamoto_M

FAIS Genome Sequencing & Analysis Group; Otomo, Y., Iida, Y.,
Fujimura, T., Ikeda, R., Ishibiki, J., Kawamata, M., Kobayashi, M.,
Kodama, T., Kurosaki, T., Kusumegi, T., Lu, M., Masuda, H., Miura, J.,
Mizuno, K., Narikawa, R., Ninkura, J., Oka, M., Ryu, R., Sugano, S.,
Sugiyama, A., Suzuki, Y., Tsunoda, Y., Ueda, M., Xie, Q., Yokomizo, S.,
Yoshimura, A., Matsubara, K. and Murakami, K.

Genome Exploration Research Group in Riken Genomic Sciences Center
and Genome Science Laboratory in Riken: Adachi, J., Aizawa, K.,
Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T.,
Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K.,
Hiraoka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y.,
Itoh, M., Kagawa, I., Kanagawa, S., Katoh, H., Kawai, J.,
Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Kouda, M.,

clone:OJ1359_D06
Published Only in Database (2002)
2 (bases 1 to 135189)
REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (20-MAR-2002) Takuji Sasaki, National Institute of
Agrobiological Sciences, Rice Genome Research Program, Kannondai
2-1-2, Tsukuba, Ibaraki 305-8602, Japan
(E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/
Tel:81-298-38-7441, Fax:81-298-38-7468)
The nucleotide sequence of this BAC clone was generated by
combining Monsanto and RGP-Japan sequencing data.
NOTE: It currently consists of 1 contigs. Gaps between the contigs
are represented as runs of N. The order of the pieces is believed
to be correct as given, however the sizes of the gaps between them
are based on estimates that have provided by the submitter. This
sequence will be replaced by the finished sequence as soon as it is
available and the accession number will be preserved.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1. 135189
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/chromosome="2"
/clone="OJ1359_D06"
ORIGIN
Query Match 24.9%; Score 130.2; DB 2; Length 135189;
Best Local Similarity 58.9%; Pred. No. 1.8e-10;
Matches 269; Conservative 0; Mismatches 173; Indels 15; Gaps 2;
QY 79 CCACCATGAGTTCTCACCACCGTTAGCCACCTCTCGACGAGACCCCTTCTCCCTTCC 138
DB 28963 CCACCTTCCCAACCTCCTCCTCCTCCTCCTCGACGACACCCACCATTTTCTTCC 29022
QY 139 CACCGCGTCT 198
DB 29023 CTCCCT 29082
QY 199 CATCTTCTGCT 258
DB 29083 GCTTCACT 29142
QY 259 CCACG-----CCGACCACTAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 306
DB 29143 ACCTTGACCT 29202
QY 307 TCACCGACCGCT 366
DB 29203 GCGATCCCT 29262
QY 367 CCACCGACGAGAGTGCACCTATGTCACCGAGTCCACCGCGCGCATGTGTCAGTGGACCA 426
DB 29263 CCCACGACGACAGTACACCTACGAGCCGCAATCCGCGCGAGGAGATCAAGTGGACCA 29322
QY 427 CCGAGACAGAGCTCTGCGCGCGCGAGAGGGTGTCTCAAGTGGAGGCTGAGCTGACTCCC 486
DB 29323 CCCACGACAGGCC--CGACGCGCGACCGGACCTTCAAGTGGAGGCGCCAGATGACACCC 29379
QY 487 CCTACGACGATGATTCGACCGGACAGTGAAGTGGA 523
DB 29380 CCAACGACGATGGCTTCGACCGCAAGTGAAGTGGA 29416
RESULT 9
LOCUS 166494 7218 bp DNA linear PAT 28-DEC-1997
DEFINITION Sequence 14 from patent US 5670367.

ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
source
1. 7218
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Query Match 20.9%; Score 109.2; DB 6; Length 7218;
Best Local Similarity 3.2%; Pred. No. 4.2e-07;
Matches 12; Conservative 260; Mismatches 98; Indels 0; Gaps 0;
QY 23 TCTCCTTACACTCCCACTCCAGCGGAGCTGAGCAGAGAGCTCCACCGCGCGGCAC 82
DB 1085 YY 1144
QY 83 CATGAGTTCTACCAACCGGTACGCCACCTCTCGACGACGCCCTTCTCCCTCCAC 142
DB 1145 YY 1204
QY 143 GCGCT 202
DB 1205 YY 1264
QY 203 CTTCGCCCTGACTCCCT 262
DB 1265 YY 1324
QY 263 GCGACCACTAGCTCCCT 322
DB 1325 YY 1384
QY 323 CCAACTGAGCTCGCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 382
DB 1385 YY 1444
QY 383 CACCTATGTC 392
DB 1445 CTCTATCTC 1454
RESULT 10
LOCUS AC022663/c 63082 bp DNA linear HTG 13-JUL-2000
AC022663
DEFINITION Homo sapiens clone RP11-2905, LOW-PASS SEQUENCE SAMPLING.
AC022663
VERSION AC022663.2 GI:9139598
KEYWORDS HTG; HTGS PHASE0.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 63082)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens, clone RP11-2905
Unpublished
REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 63082)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Dearellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagoes,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

TITLE
JOURNAL

COMMENT

Landers, T., Lehoczký, J., Levine, R., Lieu, C., Liu, G., Locke, K.,
Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
McPheeters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,
Norman, C.H., O'Connor, T., O'Donnell, P., Olivar, T.M., Peterson, K.,
Pierre, N., Pisaní, C., Pollara, V., Raymond, C., Riley, R., Rochman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange-Thomann, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirrell, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W.J.,
Zimmer, A. and Zody, M.

Direct Submission

Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 13, 2000 this sequence version replaced gi:6910643.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center Project name: L4854

Center Clone name: 29_O_5

* NOTE: This record contains 77 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 735: contig of 734 bp in length
* 735 834: gap of 100 bp
* 835 1556: contig of 722 bp in length
* 1557 1656: gap of 100 bp
* 1657 2352: contig of 696 bp in length
* 2353 2452: gap of 100 bp
* 2453 3148: contig of 696 bp in length
* 3149 3248: gap of 100 bp
* 3249 3945: contig of 697 bp in length
* 3946 4045: gap of 100 bp
* 4046 4728: contig of 683 bp in length
* 4729 4828: gap of 100 bp
* 4829 5548: contig of 720 bp in length
* 5549 5648: gap of 100 bp
* 5649 6362: contig of 714 bp in length
* 6363 6462: gap of 100 bp
* 6463 7171: contig of 709 bp in length
* 7172 7271: gap of 100 bp
* 7272 7997: contig of 726 bp in length
* 7998 8097: gap of 100 bp
* 8098 8835: contig of 738 bp in length
* 8836 8935: gap of 100 bp
* 8936 9664: contig of 729 bp in length
* 9665 9764: gap of 100 bp
* 9765 10515: contig of 751 bp in length
* 10516 10615: gap of 100 bp
* 10616 11429: contig of 814 bp in length
* 11430 11529: gap of 100 bp
* 11530 12294: contig of 765 bp in length
* 12295 12394: gap of 100 bp
* 12395 13176: contig of 782 bp in length
* 13177 13276: gap of 100 bp
* 13277 13991: contig of 715 bp in length
* 13992 14091: gap of 100 bp
* 14092 14797: contig of 706 bp in length
* 14798 14897: gap of 100 bp
* 14898 15607: contig of 710 bp in length

* 15608 15708: gap of 100 bp
* 15708 16432: contig of 725 bp in length
* 16433 16532: gap of 100 bp
* 16533 17268: contig of 736 bp in length
* 17269 17368: gap of 100 bp
* 17369 18105: contig of 737 bp in length
* 18106 18205: gap of 100 bp
* 18206 18888: contig of 683 bp in length
* 18889 18988: gap of 100 bp
* 18989 19702: contig of 714 bp in length
* 19703 19802: gap of 100 bp
* 19803 20516: contig of 714 bp in length
* 20517 20616: gap of 100 bp
* 20617 21330: contig of 714 bp in length
* 21331 21430: gap of 100 bp
* 21431 22143: contig of 713 bp in length
* 22144 22243: gap of 100 bp
* 22244 22950: contig of 707 bp in length
* 22951 23050: gap of 100 bp
* 23051 23785: contig of 735 bp in length
* 23786 23885: gap of 100 bp
* 23886 24606: contig of 721 bp in length
* 24607 24706: gap of 100 bp
* 24707 25447: contig of 741 bp in length
* 25448 25547: gap of 100 bp
* 25548 26276: contig of 729 bp in length
* 26277 26376: gap of 100 bp
* 26377 27107: contig of 731 bp in length
* 27108 27207: gap of 100 bp
* 27208 27926: contig of 719 bp in length
* 27927 28026: gap of 100 bp
* 28027 28737: contig of 711 bp in length
* 28738 28837: gap of 100 bp
* 28838 29559: contig of 722 bp in length
* 29560 29660: gap of 100 bp
* 29660 30374: contig of 715 bp in length
* 30375 30474: gap of 100 bp
* 30475 31184: contig of 710 bp in length
* 31185 31284: gap of 100 bp
* 31285 32011: contig of 727 bp in length
* 32012 32111: gap of 100 bp
* 32112 32836: contig of 725 bp in length
* 32837 32936: gap of 100 bp
* 32937 33621: contig of 685 bp in length
* 33622 33721: gap of 100 bp
* 33722 34444: contig of 723 bp in length
* 34445 34544: gap of 100 bp
* 34545 35253: contig of 709 bp in length
* 35254 35353: gap of 100 bp
* 35354 36069: contig of 716 bp in length
* 36070 36169: gap of 100 bp
* 36170 36878: contig of 709 bp in length
* 36879 36978: gap of 100 bp
* 36979 37677: contig of 699 bp in length
* 37678 37777: gap of 100 bp
* 37778 38489: contig of 712 bp in length
* 38490 38589: gap of 100 bp
* 38590 39297: contig of 708 bp in length
* 39298 39397: gap of 100 bp
* 39398 40125: contig of 728 bp in length
* 40126 40225: gap of 100 bp
* 40226 40945: contig of 720 bp in length
* 40946 41045: gap of 100 bp
* 41046 41767: contig of 722 bp in length
* 41768 41867: gap of 100 bp
* 41868 42603: contig of 736 bp in length
* 42604 42703: gap of 100 bp
* 42704 43420: contig of 717 bp in length
* 43421 43520: gap of 100 bp
* 43521 44223: contig of 703 bp in length
* 44224 44323: gap of 100 bp
* 44324 45041: contig of 718 bp in length
* 45042 45141: gap of 100 bp

* 45142 45855: contig of 714 bp in length
* 45856 45955: gap of 100 bp
* 45956 46679: contig of 724 bp in length
* 46680 46779: gap of 100 bp
* 46780 47482: contig of 703 bp in length
* 47483 47582: gap of 100 bp
* 47583 48317: contig of 735 bp in length
* 48318 48417: gap of 100 bp
* 48418 49147: contig of 730 bp in length
* 49148 49247: gap of 100 bp
* 49248 49967: contig of 720 bp in length
* 49968 50067: gap of 100 bp
* 50068 50772: contig of 705 bp in length
* 50773 50872: gap of 100 bp
* 50873 51568: contig of 696 bp in length
* 51569 51668: gap of 100 bp
* 51669 52399: contig of 731 bp in length
* 52400 52499: gap of 100 bp
* 52500 53224: contig of 725 bp in length
* 53225 53324: gap of 100 bp
* 53325 54013: contig of 689 bp in length
* 54014 54113: gap of 100 bp
* 54114 54844: contig of 731 bp in length
* 54845 54944: gap of 100 bp
* 54945 55660: contig of 716 bp in length
* 55661 55760: gap of 100 bp
* 55761 56490: contig of 730 bp in length
* 56491 56590: gap of 100 bp
* 56591 57333: contig of 743 bp in length
* 57334 57433: gap of 100 bp
* 57434 58161: contig of 728 bp in length
* 58162 58261: gap of 100 bp

Query Match 20.7%; Score 108.2; DB 2; Length 63082;
Best Local Similarity 47.6%; Pred. No. 3.5e-07;
Matches 221; Conservative 0; Mismatches 243; Indels 0; Gaps 0;

QY 16 CACCAAGTCTCTCTTACACTCCCACTCCGAGCGGAGCTGAGCAGAGAGCTCCACCGCCGC 75
Db 36038 CCCCCACCCCCCCCCCCCCCCCCCCCCCCCCACCCCCCCCCCCCCCCCCCCCCCCC 35979
QY 76 CGGCCACCATGAGTTCTCACCACCGTTACGCCACCTCTCGACGACCCCTTCTCCCT 135
Db 35978 CC 35919
QY 136 TCCCAACCGCGTCT 195
Db 35918 CC 35859
QY 196 GCGCATTTCTGGCCCTGACTCCCCCTTGCTGCTGACTGACCTTTTCATCTCCACCCCTTCC 255
Db 35858 CC 35799
QY 256 TCCCAACCGCGACTAGCT 315
Db 35798 CC 35739
QY 316 GCGTCTCCCACTGAGCTGCGCCCTGCGCGCCCGCCGACACCCCAAGGCCACGACA 375
Db 35738 CC 35679
QY 376 GGAAGTGACCTATGTACCGAGTCCACCGGAGCCGATGTCTCAAGTGAGACCAACGAGACA 435
Db 35678 CCCNNNNNNCCNNNNNNCCNNNNNNCCNNNNNNCCNNNNNNNNNNNNNNNNNNNNNN 35619
QY 436 AGCTTCGCGCGCGAGAGGGTCTCAAGTGGAGGCTGAGCTC 479
Db 35618 ANCTTCACCTGTGTGCGCGCCCTCCAGGCTCTGTGTCTC 35575

RESULT 11
AC084799/c 303091 bp DNA linear HTG 17-NOV-2000
LOCUS AC084799

DEFINITION Mus musculus chromosome 16 clone RP23-197M9, WORKING DRAFT
SEQUENCE, 101 unordered pieces.
ACCESSION AC084799
VERSION AC084799.1 GI:1192127
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 303091)
DOE Joint Genome Institute.
Sequencing of Human Chromosome 16
Unpublished
2 (bases 1 to 303091)
DOE Joint Genome Institute.
Direct Submission
Submitted (17-NOV-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 0
Center clone name: RPCI-23_197M9

Summary Statistics
Consensus quality: 152568 bases at least Q40
Consensus quality: 175579 bases at least Q30
Consensus quality: 186949 bases at least Q20
Estimated insert size: 198300; agarose-fp estimation
Estimated insert size: 293091; sum-of-contigs estimation
Quality coverage: 4.85 in Q20 bases; agarose-fp estimation
Quality coverage: 3.28 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 101 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1116: contig of 1116 bp in length
* 1117 1216: gap of unknown length
* 1217 2536: contig of 1320 bp in length
* 2537 2636: gap of unknown length
* 2637 4057: contig of 1421 bp in length
* 4058 4157: gap of unknown length
* 4158 5245: contig of 1088 bp in length
* 5246 5345: gap of unknown length
* 5346 6461: contig of 1116 bp in length
* 6462 6561: gap of unknown length
* 6562 7725: contig of 1164 bp in length
* 7726 7825: gap of unknown length
* 7826 9143: contig of 1318 bp in length
* 9144 9243: gap of unknown length
* 9244 10492: contig of 1249 bp in length
* 10493 10592: gap of unknown length
* 10593 11666: contig of 1074 bp in length
* 11667 11766: gap of unknown length
* 11767 12950: contig of 1184 bp in length
* 12951 13050: gap of unknown length
* 13051 14285: contig of 1235 bp in length
* 14286 14385: gap of unknown length
* 14386 15478: contig of 1093 bp in length
* 15479 15578: gap of unknown length
* 15579 16779: contig of 1201 bp in length
* 16780 16879: gap of unknown length
* 16880 18278: contig of 1399 bp in length
* 18279 18379: gap of unknown length
* 18379 19452: contig of 1074 bp in length
* 19453 19552: gap of unknown length

*	19553	21012:	contig of 1460 bp in length
*	21013	21112:	gap of unknown length
*	21113	22169:	contig of 1057 bp in length
*	22170	22367:	gap of unknown length
*	22270	23367:	contig of 1098 bp in length
*	22368	23467:	gap of unknown length
*	23468	24559:	contig of 1092 bp in length
*	24560	24659:	gap of unknown length
*	24660	25825:	contig of 1166 bp in length
*	25826	25925:	gap of unknown length
*	25926	27036:	contig of 1111 bp in length
*	27037	27136:	gap of unknown length
*	27137	28345:	contig of 1109 bp in length
*	28246	28345:	gap of unknown length
*	28346	29656:	contig of 1311 bp in length
*	29657	29756:	gap of unknown length
*	29757	30883:	contig of 1127 bp in length
*	30884	30983:	gap of unknown length
*	30984	32125:	contig of 1142 bp in length
*	32126	32225:	gap of unknown length
*	32226	33397:	contig of 1172 bp in length
*	33398	33497:	gap of unknown length
*	33498	34645:	contig of 1148 bp in length
*	34646	34745:	gap of unknown length
*	34746	35694:	contig of 1149 bp in length
*	35695	35994:	gap of unknown length
*	35995	37156:	contig of 1162 bp in length
*	37157	37256:	gap of unknown length
*	37257	38396:	contig of 1140 bp in length
*	38397	38496:	gap of unknown length
*	38497	39570:	contig of 1074 bp in length
*	39571	39670:	gap of unknown length
*	39671	40852:	contig of 1182 bp in length
*	40853	40952:	gap of unknown length
*	40953	42116:	contig of 1164 bp in length
*	42117	42216:	gap of unknown length
*	42217	43372:	contig of 1156 bp in length
*	43373	43472:	gap of unknown length
*	43473	44603:	contig of 1131 bp in length
*	44604	44703:	gap of unknown length
*	44704	45924:	contig of 1221 bp in length
*	45925	46024:	gap of unknown length
*	46025	47174:	contig of 1150 bp in length
*	47175	47274:	gap of unknown length
*	47275	48744:	contig of 1470 bp in length
*	48745	48844:	gap of unknown length
*	48845	49969:	contig of 1125 bp in length
*	49970	50069:	gap of unknown length
*	50070	51135:	contig of 1066 bp in length
*	51136	51235:	gap of unknown length
*	51236	52690:	contig of 1455 bp in length
*	52691	52790:	gap of unknown length
*	52791	53920:	contig of 1130 bp in length
*	53921	54020:	gap of unknown length
*	54021	55101:	contig of 1081 bp in length
*	55102	55201:	gap of unknown length
*	55202	56324:	contig of 1123 bp in length
*	56325	56424:	gap of unknown length
*	56425	57583:	contig of 1159 bp in length
*	57584	57683:	gap of unknown length
*	57684	58822:	contig of 1139 bp in length
*	58823	58922:	gap of unknown length
*	58923	60547:	contig of 1625 bp in length
*	60548	60647:	gap of unknown length
*	60648	61743:	contig of 1096 bp in length
*	61744	61843:	gap of unknown length
*	61844	62998:	contig of 1155 bp in length
*	62999	63098:	gap of unknown length
*	63099	64220:	contig of 1122 bp in length
*	64221	64320:	gap of unknown length
*	64321	65398:	contig of 1078 bp in length
*	65399	65498:	gap of unknown length
*	65499	66762:	contig of 1264 bp in length

*	66763	66862:	gap of unknown length
*	66863	68025:	contig of 1163 bp in length
*	68026	68125:	gap of unknown length
*	68126	69400:	contig of 1275 bp in length
*	69401	69500:	gap of unknown length
*	69501	70718:	contig of 1218 bp in length
*	70719	70818:	gap of unknown length
*	70819	71995:	contig of 1177 bp in length
*	71996	72095:	gap of unknown length
*	72096	73232:	contig of 1137 bp in length
*	73233	73332:	gap of unknown length
*	73333	74476:	contig of 1144 bp in length
*	74477	74576:	gap of unknown length
*	74577	75863:	contig of 1287 bp in length
*	75864	75963:	gap of unknown length
*	75964	77021:	contig of 1058 bp in length
*	77022	77121:	gap of unknown length
*	77122	78312:	contig of 1091 bp in length
*	78213	78312:	gap of unknown length
*	78313	79424:	contig of 1112 bp in length
*	79425	79524:	gap of unknown length
*	79525	80617:	contig of 1093 bp in length
*	80618	80717:	gap of unknown length
*	80718	81804:	contig of 1087 bp in length
*	81805	81904:	gap of unknown length
*	81905	83007:	contig of 1103 bp in length
*	83008	83107:	gap of unknown length
*	83108	84167:	contig of 1060 bp in length
*	84168	84267:	gap of unknown length
*	84268	85347:	contig of 1080 bp in length
*	85348	85447:	gap of unknown length
*	85448	86521:	contig of 1074 bp in length
*	86522	86621:	gap of unknown length
*	86622	87744:	contig of 1123 bp in length
*	87745	87844:	gap of unknown length
*	87845	88965:	contig of 1121 bp in length
*	88966	89065:	gap of unknown length
*	89066	90176:	contig of 1111 bp in length
*	90177	90276:	gap of unknown length
*	90277	91279:	contig of 1003 bp in length
*	91280	91379:	gap of unknown length
*	91380	92563:	contig of 1184 bp in length
*	92564	92663:	gap of unknown length
*	92664	93688:	contig of 1025 bp in length
*	93689	93788:	gap of unknown length
*	93789	94934:	contig of 1146 bp in length
*	94935	95034:	gap of unknown length
*	95035	96111:	contig of 1077 bp in length
*	96112	96211:	gap of unknown length
*	96212	98352:	contig of 2141 bp in length
*	98353	98452:	gap of unknown length
*	98453	98455:	contig of 1193 bp in length
*	99646	99745:	gap of unknown length

Query Match	20.5%;	Score 107;	DB 2;	Length 303091;
Best Local Similarity	51.9%;	Pred. No. 3.7e-07;		
Matches 182;	Conservative	0;	Mismatches 169;	Indels 0;
				Gaps 0;

OY	24	CTCCTTCACTATCCCACTCCAGCGGACGTGAGCAGAGAGTGATCCACC	CGCGCGGGCCACC	83
Db	44304	CCCCNCNCCCCCCCCCNCCCCCNCCCCCNCCCCCNCCCCCNCCCC	CCCCCNCCC	44245
OY	84	ATGAGTTTCACCAACCGTTACGCCCACTCTCGACGACCCCTTCTTCC	CCTTCCGACG	143
Db	44244	CCCNCCCCCCCCCCCCCNCCCNCCCCCNCCCCCNCCCCCCCCCCCC	CCCNCCCC	44185
OY	144	CGGTCTCTCTCTCTGCCCCCTCTCTCTCCCAACCGCGGTTTCTTCCA	CTGCATTTC	203
Db	44184	CCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCNCCNCCCCCCCC	CCCMCNC	44125
OY	204	TTCGCCCTCGACTCCCGCCTTCGCTGTAACCCCTTTCATCTCCACCC	CTTCTCCGACG	263
Db	44124	CCCCCCCCCCCCCCCCCNCCNCCCCCCCCCCCCCNCCNCCCCCCCC	CCCN	44065

QY	264	CGGACCTAGCTCCCTCTCCTCGACCCCTTCTCTCCACACACCCCTACCGACCGGTCTCC	323
Db	44064	NCCCCNCCNCCNCCCCCNCNCC	44005
QY	324	CAACTCGAGCTCGCCCTTGCCTCGCCGCGCGCCGACACCCAGGCCACGACG	374
Db	44004	CCNCCCNCCNCCNCCCCCCCCCNCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC	43954
RESULT 12			
AC087228/c		265537 bp	DNA linear HTG 19-DEC-2000
LOCUS			
DEFINITION		Mus musculus chromosome 16 clone RP23-125D15, WORKING DRAFT	
ACCESSION		AC087228	
VERSION		AC087228.1	GI:11890790
KEYWORDS		HTG; HTGS_PHASE1; HTGS_DRAFT.	
SOURCE		Mus musculus (house mouse)	
ORGANISM		Mus musculus	
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS		1 (bases 1 to 265537)	
JOURNAL		DOE Joint Genome Institute.	
REFERENCE		Sequencing of Human Chromosome 16	
AUTHORS		2 (bases 1 to 265537)	
TITLE		Unpublished	
REFERENCE		DOE Joint Genome Institute.	
AUTHORS		Submitted (19-DEC-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
JOURNAL		-----Genome Center	
COMMENT		Center: Joint Genome Institute	
		Center Code: JGI	
		Web site: http://www.jgi.doe.gov	

		Project Information	
		Center Project Name: 0	
		Center clone name: RPCI-23_125D15	

		Summary Statistics	
		Consensus quality: 162494 bases at least Q40	
		Consensus quality: 190921 bases at least Q30	
		Consensus quality: 204528 bases at least Q20	
		Estimated insert size: 249300; agarose-fp estimation	
		Estimated insert size: 260337; sum-of-contigs estimation	
		Quality coverage: 3.18 in Q20 bases; agarose-fp estimation	
		Quality coverage: 3.04 in Q20 bases; sum-of-contigs estimation	
		* NOTE: This is a 'working draft' sequence. It currently	
		* consists of 53 contigs. The true order of the pieces	
		* is not known and their order in this sequence record is	
		* arbitrary. Gaps between the contigs are represented as	
		* runs of N, but the exact sizes of the gaps are unknown.	
		* This record will be updated with the finished sequence	
		* as soon as it is available and the accession number will	
		* be preserved.	
		1 1359: contig of 1359 bp in length	
		* 1360 1459: gap of unknown length	
		* 1460 2660: contig of 1201 bp in length	
		* 2661 2760: gap of unknown length	
		* 2761 4056: contig of 1296 bp in length	
		* 4057 4156: gap of unknown length	
		* 4157 5449: contig of 1293 bp in length	
		* 5450 5549: gap of unknown length	
		* 5550 6845: contig of 1296 bp in length	
		* 6846 6945: gap of unknown length	
		* 6946 8372: contig of 1427 bp in length	
		* 8373 8472: gap of unknown length	
		* 8473 9614: contig of 1142 bp in length	
		* 9615 9714: gap of unknown length	
		* 9715 10804: contig of 1090 bp in length	
		* 10805 10904: gap of unknown length	
		* 10905 12111: contig of 1207 bp in length	

*	12112	12211:	gap of	unknown	length
*	12212	13433:	contig of	1222	bp in length
*	13434	13533:	gap of	unknown	length
*	13534	14784:	contig of	1251	bp in length
*	14785	14884:	gap of	unknown	length
*	14885	16183:	contig of	1299	bp in length
*	16184	16283:	gap of	unknown	length
*	16284	17523:	contig of	1240	bp in length
*	17524	17623:	gap of	unknown	length
*	17624	18729:	contig of	1106	bp in length
*	18730	18829:	gap of	unknown	length
*	18830	19912:	contig of	1083	bp in length
*	19913	20012:	gap of	unknown	length
*	20013	21385:	contig of	1373	bp in length
*	21386	21485:	gap of	unknown	length
*	21486	22584:	contig of	1099	bp in length
*	22585	22684:	gap of	unknown	length
*	22685	23841:	contig of	1157	bp in length
*	23842	23941:	gap of	unknown	length
*	23942	25076:	contig of	1135	bp in length
*	25077	25176:	gap of	unknown	length
*	25177	26254:	contig of	1078	bp in length
*	26255	26354:	gap of	unknown	length
*	26355	27414:	contig of	1060	bp in length
*	27415	27514:	gap of	unknown	length
*	27515	28615:	contig of	1101	bp in length
*	28616	28715:	gap of	unknown	length
*	28716	29802:	contig of	1087	bp in length
*	29803	29902:	gap of	unknown	length
*	29903	31019:	contig of	1117	bp in length
*	31020	31119:	gap of	unknown	length
*	31120	32286:	contig of	1167	bp in length
*	32287	32386:	gap of	unknown	length
*	32387	33417:	contig of	1031	bp in length
*	33418	33517:	gap of	unknown	length
*	33518	34716:	contig of	1199	bp in length
*	34717	34816:	gap of	unknown	length
*	34817	35685:	contig of	1069	bp in length
*	35686	35985:	gap of	unknown	length
*	35986	37364:	contig of	1379	bp in length
*	37365	37464:	gap of	unknown	length
*	37465	39024:	contig of	1560	bp in length
*	39025	39124:	gap of	unknown	length
*	39125	40579:	contig of	1455	bp in length
*	40580	40679:	gap of	unknown	length
*	40680	42777:	contig of	2098	bp in length
*	42778	42877:	gap of	unknown	length
*	42878	45468:	contig of	2591	bp in length
*	45469	45568:	gap of	unknown	length
*	45569	48759:	contig of	3191	bp in length
*	48760	48859:	gap of	unknown	length
*	48860	51099:	contig of	2240	bp in length
*	51100	51199:	gap of	unknown	length
*	51200	52748:	contig of	1549	bp in length
*	52749	52848:	gap of	unknown	length
*	52849	57306:	contig of	4458	bp in length
*	57307	57406:	gap of	unknown	length
*	57407	61423:	contig of	4017	bp in length
*	61424	61523:	gap of	unknown	length
*	61524	70962:	contig of	9439	bp in length
*	70963	71062:	gap of	unknown	length
*	71063	79404:	contig of	8342	bp in length
*	79405	79504:	gap of	unknown	length
*	79505	87660:	contig of	8156	bp in length
*	87661	87760:	gap of	unknown	length
*	87761	98601:	contig of	10841	bp in length
*	98602	98701:	gap of	unknown	length
*	98702	107491:	contig of	8790	bp in length
*	107492	107591:	gap of	unknown	length
*	107592	116512:	contig of	8921	bp in length
*	116513	116612:	gap of	unknown	length
*	116613	127495:	contig of	10883	bp in length
*	127496	127595:	gap of	unknown	length

*	9670	10383:	contig of 714 bp	in length
*	10384	10483:	gap of 100 bp	
*	10484	11173:	contig of 690 bp	in length
*	11174	11173:	gap of 100 bp	
*	11274	11981:	contig of 708 bp	in length
*	11982	12081:	gap of 100 bp	
*	12082	12798:	contig of 717 bp	in length
*	12899	12898:	gap of 100 bp	
*	12899	13612:	contig of 714 bp	in length
*	13613	13712:	gap of 100 bp	
*	13713	14417:	contig of 705 bp	in length
*	14418	14517:	gap of 100 bp	
*	14518	15219:	contig of 702 bp	in length
*	15220	15319:	gap of 100 bp	
*	15320	16013:	contig of 694 bp	in length
*	16014	16113:	gap of 100 bp	
*	16114	16809:	contig of 696 bp	in length
*	16810	16909:	gap of 100 bp	
*	16910	17625:	contig of 716 bp	in length
*	17626	17725:	gap of 100 bp	
*	17726	18429:	contig of 704 bp	in length
*	18430	18529:	gap of 100 bp	
*	18530	19250:	contig of 721 bp	in length
*	19251	19350:	gap of 100 bp	
*	19351	20048:	contig of 698 bp	in length
*	20049	20148:	gap of 100 bp	
*	20149	20853:	contig of 705 bp	in length
*	20854	20953:	gap of 100 bp	
*	20954	21655:	contig of 702 bp	in length
*	21656	21755:	gap of 100 bp	
*	21756	22464:	contig of 709 bp	in length
*	22465	22564:	gap of 100 bp	
*	22565	23255:	contig of 691 bp	in length
*	23256	23355:	gap of 100 bp	
*	23356	24062:	contig of 707 bp	in length
*	24063	24162:	gap of 100 bp	
*	24163	24871:	contig of 709 bp	in length
*	24872	24971:	gap of 100 bp	
*	24972	25649:	contig of 678 bp	in length
*	25650	25749:	gap of 100 bp	
*	25750	26452:	contig of 703 bp	in length
*	26453	26552:	gap of 100 bp	
*	26553	27274:	contig of 722 bp	in length
*	27275	27374:	gap of 100 bp	
*	27375	28083:	contig of 709 bp	in length
*	28084	28183:	gap of 100 bp	
*	28184	28873:	contig of 690 bp	in length
*	28874	28973:	gap of 100 bp	
*	28974	29663:	contig of 690 bp	in length
*	29664	29763:	gap of 100 bp	
*	29764	30495:	contig of 732 bp	in length
*	30496	30595:	gap of 100 bp	
*	30596	31309:	contig of 714 bp	in length
*	31310	31409:	gap of 100 bp	
*	31410	32114:	contig of 705 bp	in length
*	32115	32214:	gap of 100 bp	
*	32215	32926:	contig of 712 bp	in length
*	32927	33026:	gap of 100 bp	
*	33027	33723:	contig of 696 bp	in length
*	33723	33822:	gap of 100 bp	
*	33823	34512:	contig of 690 bp	in length
*	34513	34612:	gap of 100 bp	
*	34613	35293:	contig of 681 bp	in length
*	35294	35393:	gap of 100 bp	
*	35394	36104:	contig of 711 bp	in length
*	36105	36204:	gap of 100 bp	
*	36205	36917:	contig of 713 bp	in length
*	36918	37017:	gap of 100 bp	
*	37018	37678:	contig of 661 bp	in length
*	37679	37778:	gap of 100 bp	
*	37779	38491:	contig of 713 bp	in length
*	38492	38591:	gap of 100 bp	
*	38592	39298:	contig of 707 bp	in length

*	39299	39398:	gap of 100 bp
*	39399	40104:	contig of 706 bp in length
*	40105	40204:	gap of 100 bp
*	40205	40922:	contig of 718 bp in length
*	40923	41022:	gap of 100 bp
*	41023	41721:	contig of 699 bp in length
*	41722	41821:	gap of 100 bp
*	41822	42504:	contig of 683 bp in length
*	42505	42604:	gap of 100 bp
*	42605	43283:	contig of 679 bp in length
*	43284	43383:	gap of 100 bp
*	43384	44112:	contig of 729 bp in length
*	44113	44212:	gap of 100 bp
*	44213	44919:	contig of 707 bp in length
*	44920	45019:	gap of 100 bp
*	45020	45710:	contig of 691 bp in length
*	45711	45810:	gap of 100 bp
*	45811	46509:	contig of 699 bp in length
*	46510	46609:	gap of 100 bp
*	46610	47226:	contig of 617 bp in length
*	47227	47326:	gap of 100 bp
*	47327	48041:	contig of 715 bp in length
*	48042	48141:	gap of 100 bp
*	48142	48845:	contig of 704 bp in length
*	48846	48945:	gap of 100 bp
*	48946	49650:	contig of 705 bp in length
*	49651	49750:	gap of 100 bp
*	49751	50448:	contig of 698 bp in length
*	50449	50548:	gap of 100 bp
*	50549	51248:	contig of 700 bp in length
*	51249	51348:	gap of 100 bp
*	51349	52028:	contig of 680 bp in length
*	52029	52128:	gap of 100 bp
*	52129	52840:	contig of 712 bp in length
*	52841	52940:	gap of 100 bp
*	52941	53713:	contig of 773 bp in length
*	53714	53813:	gap of 100 bp
*	53814	54528:	contig of 715 bp in length
*	54529	54628:	gap of 100 bp
*	54629	55327:	contig of 699 bp in length
*	55328	55427:	gap of 100 bp

Query Match	20.3%	Score 106;	DB 2;	Length 69722;
Best Local Similarity	42.2%;	Pred. No. 7.3e-07;		
Matches 199; Conservative	0;	Mismatches 273;	Indels 0;	Gaps 0;

[illegible]

QY 452 GAGGTGCTCAAGTGGAGAGCTGACTGACCTCCCTAGACAGATGATTTC 503
Db 7253 NNNNNNNNNNNNNNNNNNNNNNGCGAGATTGATTATCGCCCTGTGTGAATTTC 7304

RESULT 14
AC079433/c
LOCUS AC079433
DEFINITION Mus musculus chromosome 16 clone RP23-76F18, WORKING DRAFT
AC079433 252689 bp DNA linear HTG 01-SEP-2000
SEQUENCE, 68 unordered pieces.
AC079433
AC079433.1 GI:9958045
VERSION HTG; HTGS PHASE1; HTGS DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 252689)
DOE Joint Genome Institute.
Sequencing of Mouse
Unpublished
2 (bases 1 to 252689)
DOE Joint Genome Institute.
Direct Submission
Submitted (01-SEP-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: <http://www.jgi.doe.gov>

Project Information
Center Project Name: 0
Center clone name: RPCI-23_76F18

Summary Statistics
Consensus quality: 136458 bases at least Q40
Consensus quality: 163599 bases at least Q30
Consensus quality: 177476 bases at least Q20
Estimated insert size: 181300; agarose-fp estimation
Quality coverage: 3.62 in Q20 bases; agarose-fp estimation
Quality coverage: 2.67 in Q20 bases; sum-of-contigs estimation
NOTE: This is a 'working draft' sequence. It currently
* consists of 68 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1137: contig of 1137 bp in length
* 1138 1237: gap of unknown length
* 1238 2277: contig of 1040 bp in length
* 2278 2377: gap of unknown length
* 2378 3385: contig of 1008 bp in length
* 3386 3485: gap of unknown length
* 3486 4560: contig of 1075 bp in length
* 4561 4660: gap of unknown length
* 4661 5675: contig of 1015 bp in length
* 5676 5775: gap of unknown length
* 5776 6965: contig of 1190 bp in length
* 6966 7065: gap of unknown length
* 7066 8274: contig of 1209 bp in length
* 8275 8374: gap of unknown length
* 8375 9464: contig of 1090 bp in length
* 9465 9564: gap of unknown length
* 9565 10789: contig of 1225 bp in length
* 10790 10889: gap of unknown length
* 10890 12400: contig of 1511 bp in length
* 12401 12500: gap of unknown length
* 12501 13800: contig of 1300 bp in length

* 13801 13900: gap of unknown length
* 13901 14942: contig of 1042 bp in length
* 14943 15042: gap of unknown length
* 15043 16173: contig of 1131 bp in length
* 16174 16273: gap of unknown length
* 16274 17387: contig of 1114 bp in length
* 17388 17487: gap of unknown length
* 17488 18621: contig of 1134 bp in length
* 18622 18721: gap of unknown length
* 18722 19865: contig of 1144 bp in length
* 19866 19965: gap of unknown length
* 19966 21154: contig of 1189 bp in length
* 21155 21254: gap of unknown length
* 21255 22423: contig of 1169 bp in length
* 22424 22523: gap of unknown length
* 22524 23578: contig of 1055 bp in length
* 23579 23678: gap of unknown length
* 23679 24843: contig of 1165 bp in length
* 24844 24943: gap of unknown length
* 24944 26121: contig of 1178 bp in length
* 26122 26221: gap of unknown length
* 26222 27281: contig of 1060 bp in length
* 27282 27381: gap of unknown length
* 27382 28476: contig of 1095 bp in length
* 28477 28576: gap of unknown length
* 28577 29790: contig of 1214 bp in length
* 29791 29891: gap of unknown length
* 29891 31045: contig of 1155 bp in length
* 31046 31145: gap of unknown length
* 31146 32282: contig of 1137 bp in length
* 32283 32382: gap of unknown length
* 32383 33632: contig of 1250 bp in length
* 33633 33732: gap of unknown length
* 33733 34843: contig of 1111 bp in length
* 34844 34943: gap of unknown length
* 34944 36150: contig of 1207 bp in length
* 36151 36250: gap of unknown length
* 36251 37448: contig of 1198 bp in length
* 37449 37548: gap of unknown length
* 37549 38670: contig of 1122 bp in length
* 38671 38770: gap of unknown length
* 38771 39895: contig of 1125 bp in length
* 39896 39995: gap of unknown length
* 39996 41166: contig of 1171 bp in length
* 41167 41266: gap of unknown length
* 41267 42375: contig of 1109 bp in length
* 42376 42475: gap of unknown length
* 42476 43590: contig of 1115 bp in length
* 43591 43690: gap of unknown length
* 43691 44884: contig of 1194 bp in length
* 44885 44984: gap of unknown length
* 44985 46320: contig of 1336 bp in length
* 46321 46420: gap of unknown length
* 46421 47501: contig of 1081 bp in length
* 47502 47601: gap of unknown length
* 47602 48760: contig of 1159 bp in length
* 48761 48860: gap of unknown length
* 48861 50518: contig of 1658 bp in length
* 50519 50618: gap of unknown length
* 50619 51650: contig of 1032 bp in length
* 51651 51750: gap of unknown length
* 51751 53115: contig of 1365 bp in length
* 53116 53215: gap of unknown length
* 53216 54309: contig of 1094 bp in length
* 54310 54409: gap of unknown length
* 54409 55507: contig of 1098 bp in length
* 55508 55607: gap of unknown length
* 55608 57306: contig of 1699 bp in length
* 57307 57406: gap of unknown length
* 57407 58553: contig of 1147 bp in length
* 58554 58653: gap of unknown length
* 58654 59880: contig of 1227 bp in length
* 59881 59980: gap of unknown length

*	59981	61015:	contig of 1035 bp in length
*	61016	61115:	gap of unknown length
*	61116	63001:	contig of 1886 bp in length
*	63002	63101:	gap of unknown length
*	63102	65183:	contig of 2082 bp in length
*	65184	65283:	gap of unknown length
*	65284	66388:	contig of 1105 bp in length
*	66389	66488:	gap of unknown length
*	66489	68174:	contig of 1686 bp in length
*	68175	68274:	gap of unknown length
*	68275	70061:	contig of 1787 bp in length
*	70062	70161:	gap of unknown length
*	70162	71207:	contig of 1046 bp in length
*	71208	71307:	gap of unknown length
*	71308	75030:	contig of 3723 bp in length
*	75031	75130:	gap of unknown length
*	75131	78766:	contig of 3636 bp in length
*	78767	78866:	gap of unknown length
*	78867	84332:	contig of 5466 bp in length
*	84333	84432:	gap of unknown length
*	84433	95132:	contig of 10700 bp in length
*	95133	95232:	gap of unknown length
*	95233	102209:	contig of 6977 bp in length
*	102210	102309:	gap of unknown length
*	102310	113650:	contig of 11341 bp in length
*	113651	113750:	gap of unknown length
*	113751	124224:	contig of 10474 bp in length
*	124225	124324:	gap of unknown length
*	124325	134966:	contig of 10642 bp in length
*	134967	135066:	gap of unknown length
*	135067	147947:	contig of 12881 bp in length
*	147948	148047:	gap of unknown length
*	148048	164664:	contig of 16617 bp in length
*	164665	164764:	gap of unknown length
*	164765	182006:	contig of 17242 bp in length
*	182007	182106:	gap of unknown length
*	182107	197897:	contig of 15791 bp in length
*	197898	197997:	gap of unknown length
*	197998	219975:	contig of 21978 bp in length
*	219976	220075:	gap of unknown length
*	220076	252689:	contig of 32614 bp in length.

```

FEATURES
source
location/Qualifiers
1..252689
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="16"
/clone="RP23-76F18"
/clone_lib="RPC1 mouse BAC library 23"

```

Query Match	20.3%;	Score 106;	DB 2;	Length 252689;
Best Local Similarity	43.3%;	Pred. No. 5.4e-07;		
Matches 214; Conservative	0;	Mismatches 280;	Indels 0;	Gaps 0;

[illegible][illegible]

```

RESULT 15
AC143602/c
LOCUS          AC143602          3281 bp    DNA
DEFINITION     Macaca mulatta clone CH250-267N12, ***
ACCESSION      AC143602
VERSION        AC143602.1  GI:29649004
KEYWORDS       HTG; HTGS_PHASE2; HTGS_PGI.
SOURCE         Macaca mulatta (rhesus_monkey)
ORGANISM       Macaca mulatta

```

REFERENCE	1 (bases 1 to 3281)
AUTHORS	Csuros, M. and Milosavljevic, A.
TITLE	Pooled genomic indexing (PGI): mathematical analysis and experiment

JOURNAL (in) Guigo, R. and Gusfield, D. (Eds.); ALGORITHMS IN BIOINFORMATICS, SECOND INTERNATIONAL WORKSHOP, WABI 2002, ROME, ITALY, SEPTEMBER 17-21, 2002, PROCEEDINGS: 10-28;

REFERENCE
AUTHORS

Milosavljevic, A., Sodergren, E., Csuros, M., Li, B., Jackson, A.R., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C., Albrooks, S.L., Amarutunge, H.C., Are, J.R., Ayele, M., Banks, T., Barbara, J., Benton, J., Bimace, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O., Demn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J., Egan, A., Eamhart, C., Edwards, C.C., Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Joliver, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratochvic, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W., Louisedge, H., Lozado, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Marondel, I., Martin, R., Martindale, A., Martinez, E., Massey, E., Mamhiney, E., McLeod, M.P., Meador, M., Mei, G., Merscher, S., Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogunu, M., Okunolu, G., Otagunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojudoakan, I., Rolfe, M.,

Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
Shooshitari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A.,
Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,
Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R.,
Washington,C., Watlington,S., Williams,G., Williams,A.,
Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
Zorilla,S., Kucherlapati,R., Weinstock,G. and Gibbs,R.

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Unpublished
3 (bases 1 to 3281)
Worley,K.C.

Direct Submission
Submitted (09-APR-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: LBSI
Center clone name: CH250-267N12

----- Summary Statistics
Chemistry: Dye-terminator Big Dye: inf% of reads
Chemistry: Dye-terminator Big Dye: inf% of reads
Consensus quality: 26 bases at least Q40
Consensus quality: 26 bases at least Q30
Consensus quality: 328 bases at least Q20

COMMENT

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
* NOTE: The contigs are based on the application
* of the PGI method using the Human genome (NCBI build 31)
* as the comparative genome:
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
1 3281: contig of 3281 bp in length.

FEATURES

source

1.3281
/organism="Macaca mulatta"
/mol_type="genomic DNA"
/db_xref="taxon:9544"
/clone="CH250-267N12"
1.3281
/note="assembly name:CH250-267N12.1B
CONFIDENCE: 0.67"

ORIGIN

Query Match 20.2%; Score 105.8; DB 2; length 3281;
Best Local Similarity 56.0%; Pred. No. 1.6e-06;
Matches 200; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 15 TCACCAAGTCTCCTTACACACTCCCACTCCAGCGGAGCTGAGCAGAGAGCTCCACCGCGG 74
DB 1776 TCCCTTGGCCCTCCCACTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1717

QY 75 CCGGCAACCATGAGTCTCAGCAACCGTTAGCGCCACCTCCTCGAGCAGACCCCTCTTCCCC 134
DB 1716 CCCCCCCCCCCCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1657

QY 135 TTCCCAACGCGCTCCTCCTCCTGCCCCCTTCTCTCTCCCAACCGCGGTTCTCTCCACC 194
DB 1656 CCCCCCCCCCCCCCCCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1597

QY 195 TGCCCATTTCTTGGCCCTGCACTCCCTCTGCTGACCTTTCCATCTCCACCCCTTC 254
DB 1596 CTCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTCCCTCCCTCCCTCCCTCCCTCC 1537

QY 255 CTCCCCAAGCGGACCACTAGCTCCCTCTGACCCCTTCTCTCCACACCTTCAACGAC 314
DB 1536 CCCCCCCCCCCCCCCCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1477

QY 315 CCGGTCTCCCACTGAGCTGCGCCCTGCGCGCCCGGCGCCCAACACCCCAAGCCACC 371
DB 1476 CCCCCGCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1420

Search completed: April 14, 2004, 15:23:04
Job time : 3348 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2004, 14:12:43 ; Search time 432 Seconds

(without alignments)
5143.073 Million cell updates/sec

Title: US-09-669-817A-4

Perfect score: 523

Sequence: 1 atctgacgacgacgaccca.....gaccgcaagtggagtgga 523

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*

1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	102	19.5	12733	6	ABK98631 Vector PE
2	102	19.5	12733	8	ACD13882 L. lactis
3	102	19.5	12739	6	ABK98592 Vector PE
4	102	19.5	12739	8	ACD13843 Plasmid p
5	101.4	19.4	3163	9	ADC87060 Human GPC
6	98.2	18.8	5452	9	ADC86736 Human GPC
7	98	18.7	320	3	AAA38186 Primer us
8	97.4	18.6	320	3	AAA38185 Primer us
9	97.2	18.6	2188	2	AAZ77506 Human ova
10	95.4	18.2	318	3	AAA38184 Primer us
11	95.2	18.2	320	3	AAA38183 Primer us
12	94.6	18.1	600	6	ABQ52497 Oligonuc1
13	94.6	18.1	600	6	ABQ52496 Oligonuc1
14	93.6	17.9	1065	6	ABT09682 Human PAL
15	92.2	17.6	1064	6	ABT09678 Human PAL
16	91.8	17.6	3133	9	ADC86738 Human GPC
17	89.4	17.1	1117	9	ADC86688 Human GPC
18	86.8	16.6	1416	7	ABZ20967 Human gen
19	85.8	16.4	1285	6	AB199656 Mouse isc
20	84.8	16.2	1000	3	AAA02484 Human col
21	83.8	16.0	1327	6	ABQ68452 Listeria
22	81.8	15.6	840	6	ABQ35494 Oligonuc1
23					

24	81.8	15.6	840	6	ABQ35495 Oligonuc1
25	81.6	15.6	10494	7	ACC70864 Oncolytic
26	81.2	15.5	922	9	ADC86708 Human GPC
27	80.8	15.4	434	4	AA184343 Human pol
28	80.8	15.4	437	7	ABX47670 Bovine ES
29	80.2	15.3	237	3	AAA81807 N. mening
30	78.6	15.0	1050	4	AAD21685 Mutaciona
31	78.6	15.0	2849	4	AAD21684 Human ret
32	78	14.9	53522	6	AAD30228 Human PKD
33	78	14.9	53526	2	AAT94101 Human PKD
34	78	14.9	53577	2	AAT18551 Human PKD
35	78	14.9	53577	2	AAT94108 Human PKD
36	77.4	14.8	615	7	ACA23975 Prokaryot
37	77.4	14.8	987	6	ABT09669 Human PAL
38	76.8	14.7	1593	3	AAA02504 Human col
39	76.6	14.6	4466	3	AAA14663 Nucleotid
40	76.6	14.6	4478	3	AAA14661 Nucleotid
41	76.6	14.6	4547	3	AAA14664 Nucleotid
42	76.6	14.6	4571	3	AAA14662 Nucleotid
43	76.6	14.6	77536	3	AAA14651 Nucleotid
44	76.4	14.6	12001	2	AAQ76213 HSV 1/ST
45	76.2	14.6	1385	6	ABQ70117 Listeria

ALIGNMENTS

RESULT 1
ID ABK98631 standard; DNA; 12733 BP.
XX
AC ABK98631;
XX
DT 21-OCT-2002 (first entry)
XX
DE Vector pPEPFL4 containing L. lactis derived promoter sequence.
XX
KW ds; promoter; gram positive bacteria; fusion promoter; T5; CP25; P32;
KW P59; P1P2; PL; xy10; tecto; trpo; malo; lambdac10; cellular proliferation;
KW antibiotic; vector.
XX
OS Lactococcus lactis.
OS Synthetic.
OS Staphylococcus aureus.
OS Leuconostoc mesenteroides.
XX
PN WO200251982-A2.
XX
PD 04-JUL-2002.
XX
PF 21-DEC-2001; 2001WO-US050250.
XX
PR 27-DEC-2000; 2000US-0259434P.
PR 06-SEP-2001; 2001US-00948993.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Wall D, Gross M;
XX
DR WPI; 2002-575374/61.
XX
PT Isolated nucleic acid comprises bacterial promoters modified to have
PT altered activity in at least one gram-positive organism, e.g. Bacillus
PT anthracis or Clostridium botulinum, useful for regulating gene expression
PT in bacteria.
XX
PS Example 3; Page 224-227; 246pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising a fusion
CC promoter comprising at least one promoter that is modified to have
CC altered activity in at least one gram-positive organism, or comprising
CC T5, CP25, P32, P59, P1P2 or PL linked to at least one operator consisting
CC of xy10, tecto, trpo, malo or lambdac10, where at least one operator is

Db 4997 CCCCCC 4992

RESULT 3

ID ABK98592/c
 XX ABK98592 standard; DNA; 12739 BP.
 AC ABK98592;
 XX
 DT 07-AUG-2003 (revised)
 DT 21-OCT-2002 (first entry)

XX Vector pPEPF1 containing XylR/XylO/CP25 sequences.

XX ds; promoter; gram positive bacteria; fusion promoter; T5; CP25; P32;
 KW P59; P1P2; PU; XylO; tetrO; trpO; malO; lambdaclO; cellular proliferation;
 KW antibiotic; vector.

XX Eubacteria.
 OS Bacteriophage lambda.
 OS Escherichia coli.
 OS Synthetic.

XX WO200251982-A2.

XX 04-JUL-2002.

XX 21-DEC-2001; 2001WO-US050250.

XX 27-DEC-2000; 2000US-0259434P.

XX 06-SEP-2001; 2001US-00948993.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Wall D, Gross M;

XX WPI; 2002-575374/61.

XX Isolated nucleic acid comprises bacterial promoters modified to have
 PT altered activity in at least one gram-positive organism, e.g. Bacillus
 PT anthracis or Clostridium botulinum, useful for regulating gene expression
 PT in bacteria.

XX Example 1; Page 206-209; 246pp; English.

XX The invention relates to an isolated nucleic acid comprising a fusion
 CC promoter comprising at least one promoter that is modified to have
 CC altered activity in at least one gram-positive organism, or comprising
 CC T5, CP25, P32, P59, P1P2 or PU linked to at least one operator consisting
 CC of xylO, tetrO, trpO, malO or lambdaclO, where at least one operator is
 CC positioned so binding of a repressor to an operator represses
 CC transcription from the fusion promoter. Also included are vectors and
 CC host cells comprising the fusion promoters, a method of identifying genes
 CC involved in cellular proliferation or required for proliferation of a
 CC prokaryotic cell using the vector, a method of identifying compounds that
 CC inhibit the proliferation of a prokaryotic cell using the vector, a
 CC method of identifying a compound that reduces the activity or level of a
 CC gene product required for proliferation of a cell using the vector, a
 CC compound identified by the methods, a method of inhibiting the activity
 CC or expression of a gene in an operon required for proliferation using the
 CC vector, manufacturing an antibiotic comprising using the vector or cell
 CC and identifying a nucleic acid with promoter activity in Enterococcus
 CC faecalis. The fusion promoters are useful for regulating nucleic acid or
 CC polypeptide expression, particularly for regulating gene expression in
 CC bacteria and for identifying proliferator-regulated genes or molecules
 CC with potential antibiotic activity. The modified promoters are also
 CC useful for replacing endogenous promoters to create cells with specific
 CC regulatable genes. The present sequence is vector (or fragment)
 CC incorporating a fusion promoter sequence of the invention. (Updated on 07
 CC -AUG-2003 to correct OS field.)
 XX
 SQ Sequence 12739 BP; 3543 A; 2572 C; 3211 G; 3413 T; 0 U; 0 Other;

Query Match 19.5%; Score 102; DB 6; Length 12739;
 Best Local Similarity 54.9%; Pred. No. 3.8e-10;
 Matches 201; Conservative 0; Mismatches 165; Indels 0; Gaps 0;

QY 6 ATCAGCAGATCACCAAGTCTCTTACACTCCCACTCCAGCGGAGCTGAGCAGAGAGCT 65
 DB 5363 ATCGGTGACACCC 5304
 QY 66 CCACCGCGCGCGGCGCCACCATGAGTTCTACACACCGGTTACGCCACCTCTCGACGACCCC 125
 DB 5303 CC 5244
 QY 126 TTCTTCCCTTCCCAACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 185
 DB 5243 CC 5184
 QY 186 TCCTTCACCTGCGCATTTCTTGAGCCCTGAGTCCCGCTTGAGTGTGACCTTTCAATCTC 245
 DB 5183 CC 5124
 QY 246 CACCCCTTCTTCCCAACCGCGGAGCTAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 305
 DB 5123 CC 5064
 QY 306 CTACCGACCGCGCTTCTTCCCACTGAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 365
 DB 5063 CC 5004
 QY 366 CCGCAC 371
 DB 5003 CCCCCC 4998

RESULT 4

ID ACD13843/c
 XX ACD13843 standard; DNA; 12739 BP.

XX ACD13843;

XX 15-AUG-2003 (first entry)

XX Plasmid pPEPF1 containing a Xyl5-CP25 fusion promoter.

XX Promoter; ds; gram positive bacteria; Staphylococcus aureus; plasmid;
 KW Enterococcus faecalis; operator; xylO; tetrO; trpO; malO; lambda-clO;
 KW cellular proliferation.

XX Lactococcus lactis.
 OS Synthetic.

XX US2003027286-A1.

XX 06-FEB-2003.

XX 21-DEC-2001; 2001US-00032393.

XX 06-SEP-2000; 2000US-0230335P.

XX 27-DEC-2000; 2000US-0259434P.

XX (HASE/) HASELBECK R.

XX (WALL/) WALL D.

XX (GROS/) GROSS M.

XX Haselbeck R, Wall D, Gross M;

XX WPI; 2003-479541/45.

XX New isolated nucleic acid comprising a fusion promoter having at least
 PT one promoter that is modified to have altered activity in at least one
 PT gram-positive organism, useful for regulating gene expression in
 PT bacteria.

XX Example 1; Page 63-68; 142pp; English.

Db	137	CC	78
QY	289	CCTTCCTCTCCACACCCCTCACCAGACCGGTCTCCCACTCGAGCTGCCCTCGCCGCC	348
Db	77	CC	18
QY	349	GCGCCCCACACCCC	362
Db	17	CCCCCCCCCCCCC	4

RESULT 8

ID AAA38185 standard; DNA; 320 BP.

AC AAA38185;

DT 15-SEP-2003 (revised)

DT 01-SEP-2000 (first entry)

DE Primer used in the analysis of a BVD genome fragment.

KM Primer; bovine viral diarrhoea virus; BVD; nucleic acid analysis;
KM diagnosis; pathological organism; detect; ss.

OS Pestivirus type 1.

PN WO2000020628-A1.

PD 13-APR-2000

PF 01-OCT-1999; 99WO-CA000915.

PR 01-OCT-1998; 98US-00165264.

PA (BIOI-) BIO-ID DIAGNOSTIC INC.

PI Vinayagamorthy T;

DR WPI; 2000-303800/26.

PT Nucleic acid analysis methods for simultaneously analyzing multiple
PT nucleic acid regions for diagnosis and differentiation of pathological
PT organisms comprises sequencing the nucleic acids in the reaction mixture.

PS Example 2; Page 23; 36pp; English.

CC This sequence represents a primer used in the analysis of a fragment of
CC the bovine viral diarrhoea virus (BVDV) genome. The primer is used to
CC illustrate the nucleic acid analysis methods of the invention. The
CC methods are used for sequencing a nucleic acid in a mixture comprising
CC two nucleic acid target sequences. The methods are used for
CC simultaneously analysing multiple nucleic acid regions in a single
CC reaction. This can allow the reliable diagnosis and differentiation of
CC pathological organisms. The methods can be adapted to use a series of
CC primers with additional sequences which allows the size of the amplified
CC region to be increased. The technique is especially useful when the usual
CC sequence of the region to be detected is known and the assay is being
CC carried out to confirm its presence e.g. to rule out a falsely positive
CC amplification reaction or to distinguish subsets of an organism of
CC interest or allelic forms of a gene associated with a disease or
CC predisposition to a disease. (Updated on 15-SEP-2003 to standardise OS
CC field)

Sequence 320 BP; 4 A; 6 C; 308 G; 2 T; 0 U; 0 Other;

Query Match	18.6%;	Score 97.4;	DB 3;	Length 320;
Best Local Similarity	57.3%;	Pred. No. 2.7e-09;		
Matches 176; Conservative	0;	Mismatches 131;	Indels 0;	Gaps 0;

65 TCCACCGCGCGCGGCCACCATGATTCTTACCAACCGTTACGCCCACTCTCTCGACGACCC 124

D b	307	TTCAGTCCC	248
Q Y	125	CTTCTCCCTTCCCACGGCGTCTCTCTCTGCCCCCTTCTCTCCCAACCGCGT	184
D b	247	CC	188
Q Y	185	TTCTCCACTGCCCATTTTCGCCCCTGACTCCCCCTTGCTGTGAACCTTTCATCT	244
D b	187	CC	128
.Q Y	245	CCACCCTTCTCCCCCACGCCGACAATACTCTCTCTGACCCCTTCTCTCACAC	304
D b	127	CC	68
Q Y	305	CCTACGACCGCGCTCTCCCACTGAGCTCGCCCTGCGCGCCGCCCCACACCCAG	364
D b	67	CC	8
Q Y	365	GCCCAAC 371	
D b	7	CCCCCCC 1	

RESULT 9
AAZ77506

ID AAZ77506 standard; cDNA; 2188 BP.

AC AAZ77506;

DT 10-APR-2000 (first entry)

DE Human ovarian tumor cDNA library derived EST fragment 57.

KM Expressed sequence tag; EST; human; ovarian tumor; anticancer;
KM gene therapy; treatment; ss.

Os Homo sapiens.

PN DE19817557-A1.

PD 21-OCT-1999

PF 09-APR-1998; 98DE-01017557.

PR 09-APR-1998; 98DE-01017557.

PA (META-) METAGEN GES GENOMFORSCHUNG MBH

PI. Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E,

DR WPI; 1999-591920/51.

PT New nucleic acid sequences expressed in ovarian, and some other, cancer
PT tissues, and derived polypeptides, for treatment of ovarian cancer and
PT identification of therapeutic agents.

PS Claim 3; Page 184-185; 310pp; German.

CC This invention describes novel nucleic acid (cDNA) sequences (A) which
CC have anticancer activity and are highly expressed in ovarian tumor tissue
CC (and some also in testis and breast cancer tissue). The products of the
CC invention can be used for gene therapy. (A) are used (i) for recombinant
CC expression of polypeptides (B) and (ii) to isolate complete genes. (B)
CC are used (i) to identify agents suitable for treatment of ovarian cancer;
CC (ii) directly for treating this form of cancer (including expression from
CC gene therapy vectors) and (iii) for generation of specific antibodies.
CC (A) are identified by assembling ESTs (expressed sequence tags) from a
CC particular tissue type before comparison of expression patterns. This
CC allows a significantly longer fragment of the gene to be revealed, so
CC should reduce the number of failures associated with the fact that ESTs
CC from different libraries may represent different parts of the same
CC unknown gene distorting the estimated frequency of occurrence in a

CC particular tissue. AAZ77450-Z77572 represent the human ovarian tumor cDNA
CC library derived EST fragments described in the method of the invention
CC and encode the protein fragments represented in AAY76505-Y76638

Sequence 2188 BP; 590 A; 567 C; 434 G; 597 T; 0 U; 0 Other;

Query Match	18.6%	Score 97.2;	DB 2;	Length 2188;
Best Local Similarity	57.1%;	Pred. No. 3e-09;		
Matches 177; Conservative	0;	Mismatches 133;	Indels 0;	Gaps 0;

QY	66	CCACCGCGGGGACCACCATGAGTTTCACCAACCGTTACGCCCACTCCTGCACGACCCC	125
Dp	5	CC	64
QY	126	TTC TTC CCTTCCCAACGCGGTCTCTCTCTCTGCCCCCTTCTCTCCCAACGCGCGTT	185
Dp	65	CC	124
QY	186	TCCTGCACCTGCGCCATTCTTGCCCTGACTCCCCCTTGCTGAGTAGCCCTTGCAATCTC	245
Dp	125	CC	184
QY	246	CACCCCTTCCTCCCAACGCGGACCACTAGCTCCCTCTCGAACCCCTTCTCTCCACACC	305
Dp	185	CC	244
QY	306	CTCACCGAAGCGGCTCTCCCACTGAGCTCGCCCTCGCGCGCGGCCCAACCCCAAG	365
Dp	245	CCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCCCTGCAACCCCAA	304
QY	366	CCCAACCAACA 375	
Dp	305	CCCTCCCCCA 314	

RESULT 10
AAA38184/C

ID AAA38184 standard; DNA; 318 BP.

AC AAA38184;

DT	15-SEP-2003	(revised)
DT	01-SEP-2000	(first entry)

DE Primer used in the analysis of a BVDV genome fragment.

KW Primer; bovine viral diarrhoea virus; BVDV; nucleic acid analysis;

KW diagnosis; pathological organism; detect; ss.

OS Pestivirus type 1.

PN WO200020628-A1

PD 13-APR-2000.

PF 01-OCT-1999; 99WO-CA000915.

PR 01-OCT-1998; 98US-00165264.

PA (BIOI-) BIO-ID DIAGNOSTIC INC.

PI Vinayagamorthy T;

DR WPI; 2000-303800/26.

PT Nucleic acid analysis methods for simultaneously analyzing multiple
PT nucleic acid regions for diagnosis and differentiation of pathological
PT organisms comprises sequencing the nucleic acids in the reaction mixture.

PS Example 2; Page 23; 36pp; English.

CC This sequence represents a primer used in the analysis of a fragment of
CC the bovine viral diarrhoea virus (BVDV) genome. The primer is used to
CC

CC illustrate the nucleic acid analysis methods of the invention. The
CC methods are used for sequencing a nucleic acid in a mixture comprising
CC two nucleic acid target sequences. The methods are used for
CC simultaneously analysing multiple nucleic acid regions in a single
CC reaction. This can allow the reliable diagnosis and differentiation of
CC pathological organisms. The methods can be adapted to use a series of
CC primers with additional sequences which allows the size of the amplified
CC region to be increased. The technique is especially useful when the usual
CC sequence of the region to be detected is known and the assay is being
CC carried out to confirm its presence e.g. to rule out a falsely positive
CC amplification reaction or to distinguish subsets of an organism of
CC interest or allelic forms of a gene associated with a disease or
CC predisposition to a disease. (Updated on 15-SEP-2003 to standardise OS
CC field)

SQ Sequence 318 BF; 3 A; 3 C; 306 G; 6 T; 0 U; 0 Other;

Query Match	18.2%;	Score 95.4;	DB 3;	Length 318;
Best Local Similarity	57.0%;	Pred. No. 6.4e-09;		
Matches 174;	Conservative 0;	Mismatches 131;	Indels 0;	Gaps 0.

QY	64	CTCCACCGCGCCGGGCGACCATGAGTTCTACCAACCGTTAGGCCCACTCTCGACGACC	123
Db	306	CACCAACC	247
QY	124	CGTTCTTCCCTTCCCAACGCGCTCTCTCTCTCTGCCCCCTTCTCTCCCAACCGCGCG	183
Db	246	CC	187
QY	184	TTTCTTCACCTGACCATTTCTTGCCCTGACTCCCCCTTGCGTGTGACCGCTTTCATC	243
Db	186	CC	127
QY	244	TCCACCCCTTCCCTCCCAAGCGGACCACTAGCTCCCTCTCGAACCCCTTCTCTCGACA	303
Db	126	CC	67
QY	304	CCCTCACCGACCGCGTCTCCCACTGAGCTCGCCCTCGCGGCGCGGCCACACCCCA	363
Db	66	CC	7
QY	364	GGCCC 368	
Db	6	CCCCC 2	

RESULT 11
AAA38183/

ID	AAA38183	standard; DNA; 320 BP.
----	----------	------------------------

AC AAA38183;

DT	15-SEP-2003	(revised)
DT	01-SEP-2000	(first entry)

DE Primer used in the analysis of a BVDV genome fragment.

KW Primer; bovine viral diarrhoea virus; BVD; nucleic acid analysis; diagnosis; pathological detection; as

diagnosis; pathological organism; detect; ss.

05 Pestivirus type 1.

PN W0200020628-A1

PD 13-APR-2000.

PF 01-OCT-1999; 99WO-CA000915.

PR 01-OCT-1998; 98US-00165264.

PA (BIOI-) BIO-ID DIAGNOSTIC INC.

PI Vinayagamorthy T;

XX WPI; 2000-303800/26.
DR

PT Nucleic acid analysis methods for simultaneously analyzing multiple
PT nucleic acid regions for diagnosis and differentiation of pathological.
PT organisms comprises sequencing the nucleic acids in the reaction mixture.

PS Example 2; Page 23; 36pp; English

CC This sequence represents a primer used in the analysis of a fragment of
CC the bovine viral diarrhoea virus (BVDV) genome. The primer is used to
CC illustrate the nucleic acid analysis methods of the invention. The
CC methods are used for sequencing a nucleic acid in a mixture comprising
CC two nucleic acid target sequences. The methods are used for
CC simultaneously analysing multiple nucleic acid regions in a single
CC reaction. This can allow the reliable diagnosis and differentiation of
CC pathological organisms. The methods can be adapted to use a series of
CC primers with additional sequences which allows the size of the amplified
CC region to be increased. The technique is especially useful when the usual
CC sequence of the region to be detected is known and the assay is being
CC carried out to confirm its presence e.g. to rule out a falsely positive
CC amplification reaction or to distinguish subsets of an organism of
CC interest or allelic forms of a gene associated with a disease or
CC predisposition to a disease. (Updated on 15-SEP-2003 to standardise OS
CC field)

SQ Sequence 320 BP; 9 A; 2 C; 305 G; 4 T; 0 U; 0 Other;

Query Match	18.2%;	Score 95.2;	DB 3;	Length 320;
Best Local Similarity	57.3%;	Pred. No. 7e-09;		
Matches 172; Conservative	0;	Mismatches 128;	Indels 0;	Gaps 0;

[illegible]

RESULT 12
ABQ52497

XX	ABQ52497;
AC	
XX	
DT	12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 39088.
XX
XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KM drug; side effect; cancer; central nervous system; cardiovascular;
KM gastrointestinal; respiratory system; single nucleotide polymorphism;
KM SNP; cell differentiation; ds.

Homo sapiens.

PN WO200218632-A2

07-MAR-2002
PD

XX	01-SEP-2001; 2001WO-EP010074.
PF	

PR	01-SEP-2000; 2000DE-01043826.
PR	05-SEP-2000; 2000DE-01044543.

PA (EPIG-) EPIGENOMICS AG.

PI Olex A, Piepenbrock C, Berlin K, Guettig D;

DR WPI; 2002-371829/40.

Determining the degree of cytosine methylation in genomic DNA, useful for PT diagnosis and prognosis, comprises selective hybridization of amplicons from chemically treated DNA.

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one member,
CC of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the
CC degree of hybridisation to both classes is determined from the label on
CC the amplicon. From the ratio of labels hybridised to the two classes of
CC oligomers, the degree of methylation is calculated. The method is used:
CC (i) for diagnosis and/or prognosis of side effects of therapeutic drugs
CC and of a wide range of diseases, e.g. cancer, disorders of the central
CC nervous, cardiovascular, gastrointestinal and respiratory systems etc.,
CC particularly by detecting mutations or single nucleotide polymorphisms
CC (SNP's); and (ii) for differentiation of cell or tissue types and for
CC investigating cell differentiation. The method allows the methylation
CC status of many C residues to be determined simultaneously. ABQ13410-
CC ABQ54121 represent genomic DNA sequences used to illustrate the method
CC for determining the degree of cytosine methylation described in the
CC disclosure of the invention

Sequence 600 BP; 44 A; 478 C; 51 G; 26 T; 0 U; 1 Other;

Query Match	18.1%;	Score 94.6;	DB 6;	Length 600;
Best Local Similarity	56.6%;	Pred. No. 9e-09;		
Matches 175; Conservative	0;	Mismatches 134;	Indels 0;	Gaps 0;

[illegible]

RESULT 13
ABQ52496/c
ID ABQ52496 standard; DNA; 600 BP.

XX		ABQ52496;	
AC		12-JUL-2002	(first entry)
DT		Oligonucleotide for detecting cytosine methylation SEQ ID NO 39087.	
DE		Human; cytosine methylation; 5'-Cpg-3'; uracil; cytosine; diagnosis;	
XX		dmg; side effect; cancer; central nervous system; cardiovascular;	
KW		gastrointestinal; respiratory system; single nucleotide polymorphism;	
XK		SNP; cell differentiation; ds.	
OS	Homo sapiens.		
XX	WO200218632-A2.		
PX	07-MAR-2002.		
PF	01-SEP-2001; 2001MO-EP010074.		
PR	01-SEP-2000; 2000DE-01043826.		
PA	05-SEP-2000; 2000DE-01044543.		
PI	(EPIG-) EPIGENOMICS AG.		
DZ	Olek A, Piepenbrock C, Berlin K, Guetig D;		
PT	WP1; 2002-371829/40.		
PS	Determining the degree of cytosine methylation in genomic DNA, useful for		
XX	diagnosis and prognosis, comprises selective hybridization of amplicons		
XX	from chemically treated DNA.		
CC	Claim 12; 56bp + Sequence Listing; 56pp; German.		
CC	This invention describes a novel method for determining the degree of		
CC	methylation of a particular cytosine in a motif 5'-Cpg-3', present in a		
CC	genomic sample of DNA. The sample is treated chemically to convert		
CC	Cytosine (C) but not methylated C, to uracil, then part of the genomic		
CC	DNA that contains the target C is amplified to form a labeled amplicon.		
CC	The amplicon is hybridised to two classes, each with at least one member,		
CC	of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the		
CC	degree of hybridisation to both classes is determined from the label on		
CC	the amplicon. From the ratio of labels hybridised to the two classes of		
CC	oligonucleotides, the degree of methylation is calculated. The method is used:		
CC	(i) for diagnosis and/or prognosis of side effects of therapeutic drugs		
CC	and of a wide range of diseases, e.g. cancer, disorders of the central		
CC	nervous, cardiovascular, gastrointestinal and respiratory systems etc.,		
CC	particularly by detecting mutations or single nucleotide polymorphisms		
CC	(SNPs); and (ii) for differentiation of cell or tissue types and for		
CC	investigating cell differentiation. The method allows the methylation		
CC	status of many C residues to be determined simultaneously. ABQ13410-		
CC	ABQ54121 represent genomic DNA sequences used to illustrate the method		
CC	for determining the degree of cytosine methylation described in the		
CC	disclosure of the invention		
SQ	Sequence 600 BP; 26 A; 51 C; 478 G; 44 T; 0 U; 1 Other;		
Query Match	18.1%; Score 94.6; DB 6; Length 600;		
Best Local Similarity	56.6% Pred. No. 9e-09;		
Matches 175; Conservative	0; Mismatches 134; Indels 0; Gaps 0		
Gy	63 GCTTCACC GCCCGGCCGCATCATTTCAACCAAGGTAAAGGGCACCTTCTGAAGAC	122	
Db	357 GCCCGCCCCCCC GGCCCCCGCCCCCGCCCCCGCCCCCGCCCCCGCCCCCGCCC	298	
Gy	123 CCCTTTCTCCCTTCCACCGCGCTCTCTCTCTCTGCCCCCTTCTTCTCCACCGCC	182	
Db	297 CCCCCCCCCCCCCCCCGCCACCCCCCCCCCCCCCGCCCCCCCCCCCCCCCCCCC	238	
Gy	183 GTTTCCTCCACTGCCCATTCTTGCCCTGACTCCCCCTTGCTGCTGACCTTTCAT	242	
Db	237 AACCCCCCCCCCCCCCCCCCCCCCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	178	

QY 243 CTCACCCCTTCTCCTCCCGACGGCCGACCACTAGCTCCCTCTCTGACCCCTTCTCTCTGAC 302
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 177 CCCCCCCCCCAGCGCCCCCGCCCCCTCCGCCCCCGCGCGCCCCCCCCCCCCCCCCCCCC 118
QY 303 ACCCTACCGACCGCGTCTTCCCACTGAGCTCGCCTCGCGCCGCGCCCCACACCC 362
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 117 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCAACCCCCCGCTCCCCCCCCCCCC 58
QY 363 AGGCCCGACC 371
| | | | |
Db 57 CCCCCCCCCC 49

RESULT 14

ID ABT096682 standard; DNA; 1065 BP.

AC ABT09682;

DT 02-DEC-2002 (first entry)

DE Human PAL-18 polynucleotide SEQ ID NO: 33.

KW Human; PAl-18; cancer; chromosome 1q41; prostate cancer; colon cancer;
KW breast cancer; cytostatic; gene; ds.

Os Homo sapiens.

PN US2002106765-A1.

PD 08-AUG-2002.

PF 12-MAR-2001; 2001US-00804682.

PR 10-MAR-2000; 2000US-0188586P.

PA (KIND/) KINDERS R J.
PA (CORE/) COREY M J.

PI Kinders RJ, Corey MJ;

DR WPI; 2002-697869/75.

PT New isolated PAI-18 polypeptide, useful for diagnosing, characterizing PT and treating disease and in determining disease susceptibility.

PS Claim 1; Page 50-51; 150pp; English.

CC The present invention relates to human PAL-18 polypeptides and
CC polynucleotides. The PAL-18 gene is found on chromosome 1q41. The
CC sequences can be used to diagnose, monitor and treat cancers,
CC particularly breast, colon and prostate cancers. The present sequence is
CC a PAL-18 polynucleotide shown in the invention

SQ Sequence 1065 BP; 98 A; 606 C; 36 G; 210 T; 0 U; 115 Other;

Query Match	17.9%;	Score 93.6;	DB 6;	Length 1065;
Best Local Similarity	48.4%;	Pred. No. 1.4e-08;		
Matches 186;	Conservative	0;	Mismatches 198;	Indels 0;
				Gaps 0;

[illegible]

QY 207 GCCCTGACTCCCCCTTGGCTGTGACCCCTTTCATCTCCACCCCTTCTCCCAAGCGG 266
 DB 488 TTCCCTCCNTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 547
 QY 267 ACCACTAGTCCCTCTCTGACCCCTTCTCTCTCCACACCTTCACGACGCGTCTCCAA 326
 DB 548 CCCCCCCCCCTCTCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 607
 QY 327 CTGAGCTGCGCTCTGCG 386
 DB 608 CCCCCCCCCCTCTCTCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 667
 QY 387 TATGTACGAGTCCACCGCGCGG 410
 DB 668 CCNCTTCCANCCCTCTCTCTC 691

RESULT 15

ABT09678
 ID ABT09678 standard; DNA; 1064 BP.

AC ABT09678;

DT 02-DEC-2002 (first entry)

DE Human PAL-18 polynucleotide SEQ ID NO: 29.

XX Human; PAL-18; cancer; chromosome 1q41; prostate cancer; colon cancer;

KW breast cancer; cytostatic; gene; ds.

XX Homo sapiens.

PN US2002106765-A1.

PD 08-AUG-2002.

PF 12-MAR-2001; 2001US-00804682.

PR 10-MAR-2000; 2000US-0188586P.

XX (KIND/) KINDERS R J.

PA (CORE/) COREY M J.

XX Kinders RJ, Corey MJ;

DR WPI; 2002-697869/75.

PT New isolated PAL-18 polypeptide, useful for diagnosing, characterizing,
 and treating disease and in determining disease susceptibility.

XX Claim 1; Page 47-48; 150pp; English.

CC The present invention relates to human PAL-18 polypeptides and
 CC polynucleotides. The PAL-18 gene is found on chromosome 1q41. The
 CC sequences can be used to diagnose, monitor and treat cancers,
 CC particularly breast, colon and prostate cancers. The present sequence is
 CC a PAL-18 polynucleotide shown in the invention

XX Sequence 1064 BP; 93 A; 601 C; 51 G; 194 T; 0 U; 125 Other;

Query Match 17.6%; Score 92.2; DB 6; Length 1064;
 Best Local Similarity 49.7%; Pred. No. 2.5e-08;
 Matches 169; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 32 CACTCCCACTCCAGCGGAGCTGAGAGAGAGCTCCACCGCGCGCGCGCGCGCGCGCG 91
 DB 472 CCCCCCCCCCG 531
 QY 92 TCACCAACGTTACGCGCACTCTCTGACGACCGCTTCTTCCCTTCCCAACGCGCTCTC 151
 DB 532 CCCCCCTCTACCGCTCTCCNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 591
 QY 152 CTCCTCTGCGCTCTCTCTCTCCCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 211

DB 592 CCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 651
 QY 212 CGACTCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 271
 DB 652 NCNCCCT 711
 QY 272 TAGCTCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 331
 DB 712 TNCNCCCGCT 771
 QY 332 GCTGCGCTGCT 371
 DB 772 TCCCCCGCT 811

Search completed: April 14, 2004, 16:15:52
 Job time : 438 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2004, 14:27:16 ; Search time 97 Seconds
(without alignments)
2992.158 Million cell updates/sec

Title: US-09-669-817A-4
Perfect score: 523
Sequence: 1 atctgacgacgacgaccca.....gaccgcaagtggagtgga 523

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	109.2	20.9	7218	1	US-08-232-463-14
2	98	18.7	320	3	US-09-165-264-14
3	97.4	18.6	320	3	US-09-165-264-13
4	97	18.5	319	3	US-09-165-264-8
5	95.8	18.3	320	3	US-09-165-264-7
6	95.4	18.2	318	3	US-09-165-264-12
7	95.2	18.2	320	3	US-09-165-264-11
8	81.2	15.5	1926	4	US-09-249-585A-4
9	81.2	15.5	1931	2	US-09-130-114-2
10	78	14.9	53526	3	US-08-658-136-2
11	78	14.9	53577	3	US-08-658-136-1
12	76.6	14.6	4466	4	US-09-410-551B-20
13	76.6	14.6	4478	4	US-09-410-551B-16
14	76.6	14.6	4547	4	US-09-410-551B-22
15	76.6	14.6	4571	4	US-09-410-551B-18
16	76.6	14.6	77536	4	US-09-410-551B-1
17	76.4	14.6	12001	1	US-08-458-568A-11
18	74.4	14.2	152331	3	US-09-128-155-16
19	73.8	14.1	1926	4	US-09-249-585A-2
20	73.8	14.1	1926	4	US-09-410-399-3
21	73.8	14.1	2580	3	US-09-050-863-2
22	73.8	14.1	2580	4	US-09-359-081-2
23	73.8	14.1	5452	2	US-09-130-114-1
24	73.8	14.1	8705	4	US-09-647-344A-14
25	73.8	14.1	9600	3	US-08-910-647-1
26	73.8	14.1	9600	4	US-09-620-925-1
27	73.8	14.1	10596	1	US-07-884-811-15

C	28	73.8	14.1	10596	1	US-07-885-971-15	Sequence 15, Appl
C	29	73.8	14.1	10596	1	US-08-087-783A-15	Sequence 15, Appl
C	30	73.8	14.1	10596	1	US-08-194-088B-15	Sequence 15, Appl
C	31	73.8	14.1	10596	2	US-08-194-087-15	Sequence 15, Appl
C	32	73.8	14.1	10596	5	PCT-US93-04648-15	Sequence 15, Appl
C	33	73.8	14.1	16080	4	US-09-724-566A-48	Sequence 48, Appl
C	34	73.6	14.1	77536	4	US-09-410-551B-1	Sequence 1, Appl
C	35	72.6	13.9	4897	6	5196516-7	Patent No. 5196516
	36	71.4	13.7	51259	3	US-08-781-891-209	Sequence 209, App
	37	71.4	13.7	51259	4	US-09-618-166-209	Sequence 209, App
	38	65.4	12.5	925	3	US-08-858-003-1	Sequence 1, Appl
	39	65.4	12.5	925	3	US-09-078-166-1	Sequence 1, Appl
	40	65.4	12.5	925	3	US-09-078-166-1	Sequence 1, Appl
	41	64.4	12.3	438	4	US-09-252-991A-8422	Sequence 8422, Ap
	42	64.4	12.3	474	4	US-09-252-991A-8385	Sequence 8385, Ap
C	43	64.4	12.3	1761	4	US-09-252-991A-8423	Sequence 8423, Ap
	44	63.8	12.2	31571	1	US-08-323-443B-1	Sequence 1, Appl
	45	63.2	12.1	4403765	3	US-09-103-840A-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-9300
; TELEFAX: (703) 683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
; US-08-232-463-14
Query Match 20.9%; Score 109.2; DB 1; Length 7218;

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-8

Query Match 18.5%; Score 97; DB 3; Length 319;
Best Local Similarity 56.9%; Pred. No. 1.8e-11;
Matches 178; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 58 AGAGAGCTCCACCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 117
DB 313 AGGAGTGTGACAGACCGCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 254
QY 118 AGGAGTGTGACAGACCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 177
DB 253 AGGAGTGTGACAGACCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 194
QY 178 CGGCGGTTTCTCCACCTGCGCATTTCTTGGCTGACTCCCGCTTGGTGTGACCGTT 237
DB 193 CGGCGGTTTCTCCACCTGCGCATTTCTTGGCTGACTCCCGCTTGGTGTGACCGTT 134
QY 238 TCCATCTCCACCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 297
DB 133 TCCATCTCCACCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 74
QY 298 TCCACACCTCTACCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 357
DB 73 TCCACACCTCTACCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 14
QY 358 ACCCGAGGCGCCAC 370
DB 13 ACCCGAGGCGCCAC 1

RESULT 5
US-09-165-264-7/c
Sequence 7, Application US/09165264
Patent No. 6197510
GENERAL INFORMATION:
APPLICANT: Vinayagamoorthy, Thuraiayah
TITLE OF INVENTION: Multi-Loci Genomic Analysis
FILE REFERENCE: 44747
CURRENT APPLICATION NUMBER: US/09/165,264
CURRENT FILING DATE: 1998-10-01
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 7
LENGTH: 320
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-7

Query Match 18.3%; Score 95.8; DB 3; Length 320;
Best Local Similarity 57.5%; Pred. No. 3.2e-11;
Matches 172; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 64 CTCGACCGCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCGAGCACC 123
DB 302 CTCGACCGCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCGAGCACC 243
QY 124 CCTTCTTCCCTTCCGACCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 183
DB 242 CCTTCTTCCCTTCCGACCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 183
QY 184 TTCTCTGACCTGCGCATTTCTTGGCTGACTCCCGCTTGGTGTGACCGTTTGCATC 243
DB 182 TTCTCTGACCTGCGCATTTCTTGGCTGACTCCCGCTTGGTGTGACCGTTTGCATC 123
QY 244 TCCAGCCCTTCTCTCCGACCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 303
DB 122 TCCAGCCCTTCTCTCCGACCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 63

QY 304 CCTGACCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCGAGCACC 362
DB 62 CCTGACCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCGAGCACC 4

RESULT 6
US-09-165-264-12/c
Sequence 12, Application US/09165264
Patent No. 6197510
GENERAL INFORMATION:
APPLICANT: Vinayagamoorthy, Thuraiayah
TITLE OF INVENTION: Multi-Loci Genomic Analysis
FILE REFERENCE: 44747
CURRENT APPLICATION NUMBER: US/09/165,264
CURRENT FILING DATE: 1998-10-01
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 12
LENGTH: 318
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-12

Query Match 18.2%; Score 95.4; DB 3; Length 318;
Best Local Similarity 57.0%; Pred. No. 3.8e-11;
Matches 174; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 64 CTCGACCGCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCGAGCACC 123
DB 306 CTCGACCGCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCGAGCACC 247
QY 124 CCTTCTTCCCTTCCGACCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 183
DB 246 CCTTCTTCCCTTCCGACCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 187
QY 184 TTCTCTGACCTGCGCATTTCTTGGCTGACTCCCGCTTGGTGTGACCGTTTGCATC 243
DB 186 TTCTCTGACCTGCGCATTTCTTGGCTGACTCCCGCTTGGTGTGACCGTTTGCATC 127
QY 244 TCCAGCCCTTCTCTCCGACCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 303
DB 126 TCCAGCCCTTCTCTCCGACCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCG 67
QY 304 CCTGACCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCGAGCACC 363
DB 66 CCTGACCGCGCGGACCATGATGTTCTACCAACCGTTAGCGCCACTCTCGAGCACC 7
QY 364 GGCCC 368
DB 6 GGCCC 2

RESULT 7
US-09-165-264-11/c
Sequence 11, Application US/09165264
Patent No. 6197510
GENERAL INFORMATION:
APPLICANT: Vinayagamoorthy, Thuraiayah
TITLE OF INVENTION: Multi-Loci Genomic Analysis
FILE REFERENCE: 44747
CURRENT APPLICATION NUMBER: US/09/165,264
CURRENT FILING DATE: 1998-10-01
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 11
LENGTH: 320
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-11


```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 4466
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9)..(4454)
US-09-410-551B-20

```

Query Match	14.6%	Score 76.6;	DB 4;	Length 4466;
Best Local Similarity	53.4%;	Pred. No. 3.2e-07;		
Matches 183; Conservative	0;	Mismatches 159;	Indels 1;	Gaps 1;

OY	32	C	A	C	T	C	C	C	A	C	T	G	A	G	C	T	C	A	C	C	G	C	C	G	C	C	A	C	C	A	T	G	A	T	T	C	91				
Db	3741	C	A	C	C	C	C	A	C	C	C	C	T	C	C	C	C	T	G	G	C	C	C	A	C	T	G	C	C	A	C	C	C	C	A	C	T	C	G	C	3800

OY 92 TCACCAACCGTACGCCCACTCTCTGAGCAGCCCCCTTTCCCTTCCCACGGCGGTCTC 151
| | | | | | | | | | | | | | | | | | | | |
DB 3801 -CTACCCACCAACCCCTCCACCAACCCCACTCACCCCTTCACAACCAACCAACCCCAAC 3859

QY	152	CTCTCTCTGCCCCCTCTCTCTCTCCCAACCGCGGCTTTCCTCCACCTGCCCATCTCTTGGCCCT	211
Db	3860	CACCAACACCCCCCTCAACCCCGAACAACGCCATCATCATCAACCGGCGCTCCGGCAACCT	3919

QY	212	CGACTCCCCCTTGCTGTGACCTTTCATCTCCACCCCTTCTCCCCACGCCGACCAC	271
Db	3920	CGCCGGCATCTCGCCGCCACTGAACACCCCCACACCTACTCTCTCCGCACCCC	3979

QY 272 TAGTCCCTCCTGACCCCCCTTCCTCTCTCCACACCCTCACCGAGCGGTCTCCCAACTCGA 331
 | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3980 ACCCCCCGACGCCACCCCCGGGCACCCACTTCCCCTGGAGCGTGCGGCGAACCCCACT 4039

Oy 332 GCTGCCCTCGCCGCGCGGGCCCCACACCAGGCCCAACCAGC 374
| | | | | | | | | | | | | |
Db 4040 CGCACCACTTCAACCCACATCCCCCAACCCCTCACCGGCATC 4082

RESULT 13
US-09-410-551B-16

Patent No. 6503737
GENERAL INFORMATION:
APPLICANT: KOSAN BIOSCIENCES, Inc.

APPLICANT: CHU, DANIEL
APPLICANT: KHOSLA, CHAITAN
APPLICANT: SANTI, DANIEL

```

; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
;
; TITLE OF INVENTION: CONSTRUCTS THEREFOR
;
; FILE REFERENCE: 30062-20026.00
;

```

```

; CURRENT FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIORITY DATE: 1999-06-17

```

; PRIOR FILING DATE: 1999-03-11
 ;
 ; PRIOR APPLICATION NUMBER: US 60/102,748
 ;
 ; PRIOR FILING DATE: 1998-10-02
 ;
 ; NUMBER OF SEQ. NO. TO WORD NO.

```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 4478
TYPE: DNA

```

```

; FEATURE:
;
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthet
; OTHER INFORMATION: PKS synthase fragment

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3)..(4466)
US-09-410-551B-16

```

Query Match	14.6%;	Score 76.6;	DB 4;	Length 4478;
Best Local Similarity	53.4%;	Pred. No. 3.2e-07;		
Matches 183; Conservative	0;	Mismatches 159;	Indels 1;	Gaps 1;

QY 32 CACTCCCACTCCAGCGGAGCTGAGCAGAGACTCCACCGCGCGGCCACCAATGATTTC 91

Db 3753 CACCCCCACACCCCCCTCCCCCTGGCCCACTGGCCACCCCTGAGACCACCCCACTCCGC 3812

QY 92 TCACCAACCGTTAGGCCCAACCTCTCGACGACCCCTTCTTCCCTTCCACAGCGCGTCTC 151
 Db 3813 -CTACCCACGACACCTTCCACCAACCCCGACCTGACCCCTTCCACACGACACCCACCC 3871

```

Oy      152 CTCTCTGCCCCCTTCTCTCTCCCCACCCCGCGTTTCTCCACCTGCCCATTTCTGGCCCT 211
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      3872 CACCACCAACCCCTCAACCCCGAACACGCCATCATCATCAGCGGCTCCGGCACTT 3931
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

QY 212 CGACTCCCCCTTGGCTGCTGACCCCTTTCCATCTCCACCCTTCCTCCCACGCCGAGCAC 271
| | | | | | | | | | | | | | | | | | | | |
Db 3932 CGCGGCATCCTGCGCCGCCACCTGAACCAACCCCAACCTACCTCCTCTCCGCAACCC 3991

Oy 272 TAGTCCCTCCTCGAACCCCCTTCCTCCTCCACAGCCTCACCGACC GGCTGTCCCAACTGA 331
| | | | |
Db 3992' ACCCCCCGACGCCACCCCCGGCACCCACCTCCCTTGCGAGCTGGCGA CCCCCCAACA ACT 4051

Qy 332 GCTGCCCTGGCGGCGCGCGCCACACCCCGAGGCCACCAGC 374
| | | | | | | | | | | | | | | | | | | | | |
Db 4052 CGCCACCACCCCTCACCCACATCCCCCAACCCCTCACCGGCATC 4094

RESULT 14
US-09-410-551B-22

Patent No. 6503737
GENERAL INFORMATION:
APPLICANT: KOSAN BIOSCIENCES, Inc.

APPLICANT: CHU, DANIEL
APPLICANT: KHOSLA, CHAITAN
APPLICANT: SANTI, DANIEL

TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
 TITLE OF INVENTION: CONSTRUCTS THEREFOR
 FILE REFERENCE: 30062-20026.00

CURRENT FILING DATE: 1999-10-01
 PRIOR APPLICATION NUMBER: US 60/139,650
 PRIOR FILING DATE: 1999-06-17

PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 60/102,748
PRIOR FILING DATE: 1998-10-02

```

; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 22
; LENGTH: 4547
;

```

```

; ORGANISM: Artificial Sequence
;
; FEATURE:
;
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
;
; SOURCE INFORMATION: DNA substrate, forward

```

```

; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9) .. (4535)

```

Query Match	14.68;	Score 76.6;	DB 4;	Length 4547;
Best Local Similarity	53.4%;	Pred. No. 3.2e-07;		
Matches 183;	Conservative 0;	Mismatches 159;	Indels 1;	Gaps 1;

QY 32 CACTCCCACTCCAGCGGAGCTGAGAGAGAGCTCCACCGCCGCCGCCACCATGAGTTC 91
Db 3822 CACCCCCACACCCCCCTCCCCCTGGGCCCACTGCGCCACCTTCGACCACCCCACTCCGC 3881
QY 92 TCACACCGTTAGCGCCACCTTCCTGACGAGACCCCTTCTTCCCTTCGCCACCGCGTCTC 151
Db 3882 -CTCACCACCAACACCTTCACACCAACCCCACTCACCCTCCACACCAACCAACCCCACTC 3940
QY 152 CTCTCTCTGCCCCCTTCTCTCTCCCAACCCCGGCTTCTCTCACTGCCCCATTTCTGCCCC 211
Db 3941 CACCAACCAACCCCTCAACCCCGAACAACGCGCATCATCAACCGCGCTCCGCAACCCCT 4000
QY 212 CGACTCCCTCTGCTGCTGACCTTTTCATCTCCACACCCCTTCTCTCCCAACCGCGCAAC 271
Db 4001 CGCGGCGATCTCTGCGCGCGCACTGAACCAACCCCACTTACCTCTCTCCGCAACCC 4060
QY 272 TAGCTCCCTCTGACCCCTTCTCTCCACACCCCTCAACCGAGCGGTCTCCCACTCGA 331
Db 4061 ACCCCCGACGCGACCCCGCGACCACTCCCTGCGAGTGGCGACCCCACTCACT 4120
QY 332 GCTGCGCTCTGCG 374
Db 4121 CGCGACCACTCTCACCACCATCCCCCAACCCCTCACCAGCATC 4163

RESULT 15

US-09-410-551B-18
; Sequence 18, Application US/09410551B
; Patent No. 6503737
; GENERAL INFORMATION:
; APPLICANT: KOSAN BIOSCIENCES, Inc.
; APPLICANT: REEVES, CHRISTOPHER
; APPLICANT: CHU, DANIEL
; APPLICANT: KHOSLA, CHAITAN
; APPLICANT: SANTI, DANIEL
; APPLICANT: WU, KAI
; TITLE OF INVENTION: POLYKETIDE SYNTHASE ENZYMES AND RECOMBINANT DNA
; TITLE OF INVENTION: CONSTRUCTS THEREFOR
; FILE REFERENCE: 30062-20026.00
; CURRENT APPLICATION NUMBER: US/09/410,551B
; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: US 60/139,650
; PRIOR FILING DATE: 1999-06-17
; PRIOR APPLICATION NUMBER: US 60/123,810
; PRIOR FILING DATE: 1999-03-11
; PRIOR APPLICATION NUMBER: US 60/102,748
; PRIOR FILING DATE: 1998-10-02
; NUMBER OF SEQ ID NOS: 72
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 4571
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: DNA encoding synthetic
; OTHER INFORMATION: PKS synthase fragment
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (9)..(4559)
US-09-410-551B-18

Query Match 14.6%; Score 76.6; DB 4; Length 4571;
Best Local Similarity 53.4%; Pred. No. 3.3e-07;
Matches 183; Conservative 0; Mismatches 159; Indels 1; Gaps 1;

QY 32 CACTCCCACTCCAGCGGAGCTGAGAGAGAGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 91
Db 3846 CACCCCAACACCCCTCTCCCTGGCGCACTGCGACCTTCGACCAACCCCACTCCGC 3905
QY 92 TCACCAACCGTTAGCG 151
Db 3906 -CTCACCACCAACCTTCACCAACCCCACTCACCCTCCCAACCAACCAACCAACCAACCAACCAAC 3964

QY 152 CTCTCTCTGCCCCCTTCTCTCTCCCAACCCCGGCTTCTCTCACTGCCCCATTTCTGCCCC 211
Db 3965 CACCAACCAACCCCTCAACCCCGAACAACGCGCATCATCAACCGCGCTCCGCAACCCCT 4024
QY 212 CGACTCCCTCTGCTGCTGACCTTTTCATCTCCACACCCCTTCTCTCCCAACCGCGCAAC 271
Db 4025 CGCGGCGATCTCTGCGCGCGCACTGAACCAACCCCACTTACCTCTCTCCGCAACCC 4084
QY 272 TAGCTCCCTCTGACCCCTTCTCTCCACACCCCTCAACCGAGCGGTCTCCCACTCGA 331
Db 4085 ACCCCCGACGCGACCCCGCGCACTCACTCCCTGCGAGTGGCGACCCCACTCACT 4144
QY 332 GCTGCGCTCTGCG 374
Db 4145 CGCGACCACTCTCACCACCATCCCCCAACCCCTCACCAGCATC 4187

Search completed: April 14, 2004, 16:19:59
Job time : 104 secs


```
Db      164 GCTTCTCTCTGACGACACCCCTTTCTACCAACCTCTGCGCTGGCATCACTCCC 223
QY      193 CCGCCCATTTCTGCGCTGACCTCCCTTGGCTGTGACCTTTTCATCTCCACCCCT 252
Db      224 GCTCCCGATGACGACACCTTGAACCTGACCTCTCTCCCTCCCTCCCAAGCCG 283
QY      253 TCCTCCCAAGCGGAC--CACTAGCTCCCTCTGACCCCTTCTCTCTCAACCTCA 309
Db      284 CCGTCTCCGCGCGCGCGCTTCTGACTTGAACCTCTCTCTCTGACACCTCTG 343
QY      310 CCGACCGGCTCTCCCACTGAGTGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCA 369
Db      344 GCCACCGCATCTCCGCTCTGAGCGCGCGCTCGC-----CCTCGCGCGCAC 388
QY      370 CCAGCAGAGTGCACCTATGTACCGAGTCCACCGCGCGCATGTCAAGTGAACCG 429
Db      389 CCGCGCGCAGTACACCTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 448
QY      430 AGCACAAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 489
Db      449 AGGACAGGCGC--CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 505
QY      490 ACGACGATGATTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 523
Db      506 ACACCGACGCGCTTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 539
```

RESULT 2

```
US-10-424-599-20495
; Sequence 20495, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 20495
; LENGTH: 815
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(815)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: clone ID: PAT_MRT3847_118511C.1
US-10-424-599-20495
```

Query Match 20.2%; Score 105.4; DB 12; Length 815;
Best Local Similarity 55.3%; Pred. No. 1.1e-16;
Matches 205; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

```
QY      1 ATCTGATCAGAGATCACCAGTCTCTTCACTCCCACTCCAGCGGAGCTGACGAGA 60
Db      280 ACCCGACCCCTATCCCAAGCCCAACCCCAACCCCAACCCCAACCCCAACCC 339
QY      61 GAGCTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
Db      340 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 399
QY      121 ACCCTTTCTTCCCTTCCCAACCGCGCTCTCTCTCTCTCTCTCTCTCTCT 180
Db      400 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 459
QY      181 CCGTTCTCTCAGCTGCGCATTTCTTGGCTTGAATCCCGCTTGGCTGACCTTTC 240
```

```
Db      460 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 519
QY      241 ATCTGACCCCTTCTCTCCCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db      520 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 579
QY      301 ACACCTTCAACGACCGCGCTTCTCACTGAGTGGCTGCGCGCGCGCGCGCGCG 360
Db      580 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 639
QY      361 CCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 371
Db      640 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 650
```

RESULT 3

```
US-10-424-599-107958/c
; Sequence 107958, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 107958
; LENGTH: 771
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(771)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: clone ID: PAT_MRT3847_68501C.1
US-10-424-599-107958
```

Query Match 19.5%; Score 102; DB 12; Length 771;
Best Local Similarity 53.3%; Pred. No. 7.6e-16;
Matches 186; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

```
QY      26 CTTCACTCTCCCACTCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 85
Db      753 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 694
QY      86 GAGTCTCAGCAGCGGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 145
Db      693 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 634
QY      146 GTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 205
Db      633 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 574
QY      206 CGCCTCTGACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 265
Db      573 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 514
QY      266 GACCACTAGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 325
Db      513 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 454
QY      326 ACTGAGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 374
Db      453 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 405
```

RESULT 4

US-10-032-393-47/c


```
; Sequence 47, Application US/10032393
; Publication No. US20030027286A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Wall, Daniel
; APPLICANT: Gross, Molly
; TITLE OF INVENTION: BACTERIAL PROMOTERS AND METHODS OF USE
; FILE REFERENCE: ELITRA.010A
; CURRENT APPLICATION NUMBER: US/10/032,393
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/259,434
; PRIOR FILING DATE: 2000-12-27
; PRIOR APPLICATION NUMBER: 09/948,993
; PRIOR FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 12733
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Vector pBPEF14
US-10-032-393-47
```

```
Query Match      19.5%; Score 102; DB 14; Length 12733;
Best Local Similarity 54.9%; Pred. No. 6.3e-16;
Matches 201; Conservative 0; Mismatches 165; Indels 0; Gaps 0;
```

```
QY      6 ATCAGCAGATCACCAGTCTCTTCTCAGACTCCCACTCCAGCGGAGCTGAGAGAGCT 65
Db      5357 ATCGTGCAGACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5298
QY      66 CCACCGCGCGCGGCCACCATGAGTTCTCACCACCGGTTAGCGCCACCTCTCGAGAGAGCC 125
Db      5297 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5238
QY      126 TTCTTCCCTTCCCAACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 185
Db      5237 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5178
QY      186 TCCTTCACCTGCGCCATCTTGCGCCCTGAGCTCCCTTGGCTGTGAGCCCTTTCATCTC 245
Db      5177 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5118
QY      246 CACCCCTTCTCTCCCAACCGCGGAGCACTAGCTCCCTCTCTGAGCCCTTCTCTCCACACC 305
Db      5117 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5058
QY      306 CTCACCGACCGCGTCTCCCACTGAGCTGCGCCCTGCGCGCGCGCGCGCGCGCGCGAGG 365
Db      5057 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 4998
QY      366 CCCACG 371
Db      4997 CCCCCC 4992
```

```
RESULT 5
US-10-032-393-8/c
; Sequence 8, Application US/10032393
; Publication No. US20030027286A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Wall, Daniel
; APPLICANT: Gross, Molly
; TITLE OF INVENTION: BACTERIAL PROMOTERS AND METHODS OF USE
; FILE REFERENCE: ELITRA.010A
; CURRENT APPLICATION NUMBER: US/10/032,393
; PRIOR FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 60/259,434
; PRIOR FILING DATE: 2000-12-27
```

```
; PRIOR APPLICATION NUMBER: 09/948,993
; PRIOR FILING DATE: 2001-09-06
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 12739
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Vector pBPEF1
US-10-032-393-8
```

```
Query Match      19.5%; Score 102; DB 14; Length 12739;
Best Local Similarity 54.9%; Pred. No. 6.3e-16;
Matches 201; Conservative 0; Mismatches 165; Indels 0; Gaps 0;
```

```
QY      6 ATCAGCAGATCACCAGTCTCTTCTCAGACTCCCACTCCAGCGGAGCTGAGAGAGCT 65
Db      5363 ATCGTGCAGACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5304
QY      66 CCACCGCGCGCGGCCACCATGAGTTCTCACCACCGGTTAGCGCCACCTCTCGAGAGAGCC 125
Db      5303 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5244
QY      126 TTCTTCCCTTCCCAACCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 185
Db      5243 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5184
QY      186 TCCTTCACCTGCGCCATCTTGCGCCCTGAGCTCCCTTGGCTGTGAGCCCTTTCATCTC 245
Db      5183 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5124
QY      246 CACCCCTTCTCTCCCAACCGCGGAGCACTAGCTCCCTCTCTGAGCCCTTCTCTCCACACC 305
Db      5123 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5064
QY      306 CTCACGACCGCGTCTCCCACTGAGCTGCGCCCTGCGCGCGCGCGCGCGCGCGCGAGG 365
Db      5063 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 5004
QY      366 CCCACG 371
Db      5003 CCCCCC 4998
```

```
RESULT 6
US-10-017-161-1857/c
; Sequence 1857, Application US/10017161
; Publication No. US20030143668A1
; GENERAL INFORMATION:
; APPLICANT: SUWA, MAKIKO
; APPLICANT: ASAI, KIYOSHI
; APPLICANT: AKIYAMA, YUTAKA
; APPLICANT: ABURATANI, HIROYUKI
; TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
; FILE REFERENCE: 084335/0152
; CURRENT APPLICATION NUMBER: US/10/017,161
; PRIOR FILING DATE: 2002-12-18
; PRIOR APPLICATION NUMBER: JP 2001/246789
; PRIOR FILING DATE: 2001-06-18
; NUMBER OF SEQ ID NOS: 2430
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1857
; LENGTH: 3163
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: source
; LOCATION: (1)..(3163)
; FEATURE:
; NAME/KEY: CDS
```

LOCATION: (201)..(2963)
FEATURE:
NAME/KEY: modified_base
LOCATION: (556)..(655)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (658)..(669)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (671)..(697)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (704)..(707)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (709)..(736)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (739)..(743)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (745)..(747)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (750)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (752)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (754)..(756)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (761)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (763)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (818)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (843)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (892)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (897)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (920)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (941)
OTHER INFORMATION: a, t, c, g, unknown or other

FEATURE:
NAME/KEY: modified_base
LOCATION: (993)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (995)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1066)..(1067)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1082)..(1084)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1105)..(1107)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1130)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1147)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1162)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1202)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1247)..(1248)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1261)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1485)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1500)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1544)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1601)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1605)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1635)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:


```

FEATURE:
NAME/KEY: modified_base
LOCATION: (843)..(843)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (892)..(892)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (897)..(897)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (920)..(920)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (941)..(941)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (993)..(993)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (995)..(995)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1066)..(1067)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1082)..(1084)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1105)..(1107)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1130)..(1130)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1147)..(1147)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1162)..(1162)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1202)..(1202)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1233)..(1233)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1247)..(1248)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1261)..(1261)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1485)..(1485)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:

```

```

NAME/KEY: modified_base
LOCATION: (1500)..(1500)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1544)..(1544)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1601)..(1601)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1605)..(1605)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1635)..(1635)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1637)..(1637)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1654)..(1654)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1674)..(1674)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1695)..(1695)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1852)..(1852)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1866)..(1866)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (1968)..(1980)
OTHER INFORMATION: a, t, c, g, unknown or other

```

Query Match	19.4%;	Score 101.4;	DB 15;	Length 3163;
Best Local Similarity	51.1%;	Pred. No. 9.7e-16;		
Matches 207;	Conservative 0;	Mismatches 198;	Indels 0;	Gaps 0;

QY 6 ATCAGCAGATACCAAGTCTCTCCTTACACTCCCACTCCAGCGGCGAGCTGAGCAGAGAGCT 65
DB 1270 AACACCCCCCNCCCGAACCCTCCCNCCCCCCCCCCCCCCCCCGGCGGCCCCCCCC 1211
QY 66 CCACCGCGCGCGGCCACCATGAGTTCTCACCACCGTTACGCCACCTCCTCGACGACCCC 125
DB 1210 CCCCCCCCNCGGCC 1151
QY 126 TTCTTCCCCCTTCCCAACGCGCGCTCTCTCTCTCTGCCCCCTTCTCTCCCCACCGCGGTT 185
DB 1150 CCCNCCCCCCCCCGCGCCCCCNCCCAACCCCCCCCCCCCCCCCCCNCCCAACCCCCCCCC 1091
QY 186 TCCTCCACCTGCCCATTTCTGCGCCTCGACTCCCCCTTGGCTGCTGACCCCTTCCATCTC 245
DB 1090 CCCCCCNNNCCCCCCCCCCCCCCCCCNCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 1031
QY 246 CACCCCTTCTCCCCACGCGGACGACACTAGTCCCTCTCGACCCCTTCTCTCCACACC 305
DB 1030 CC 971
QY 306 CTCACCGACCGCGTCTTCCCACTCGAGCTGCGCCTGCGCGCGCGCGCGGCCACACCCCAAG 365


```
LOCATION: (162)..(162)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (165)..(165)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (168)..(168)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (172)..(172)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (189)..(189)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (191)..(193)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (196)..(196)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (199)..(199)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (214)..(214)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (218)..(218)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (221)..(221)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (225)..(226)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (229)..(229)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (232)..(232)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (235)..(236)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (246)..(246)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (248)..(252)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (256)..(257)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (261)..(263)
OTHER INFORMATION: a, t, c, g, unknown or other

OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (265)..(266)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (271)..(273)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (276)..(277)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (283)..(285)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (296)..(296)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (298)..(300)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (302)..(302)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (307)..(307)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (314)..(315)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (334)..(334)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (339)..(339)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (343)..(343)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (345)..(348)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (351)..(352)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (357)..(357)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (359)..(359)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (362)..(362)
OTHER INFORMATION: a, t, c, g, unknown or other
FEATURE:
NAME/KEY: modified_base
LOCATION: (365)..(365)
OTHER INFORMATION: a, t, c, g, unknown or other
```


D_b 541 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCGNCCNCCCCCCCCCCCCCCCACCC 482

QY 271 CTAGTCCTCCTCGACCCCTTCTCTCCACACCCTCACGGCGGTGCCAACTCG 330
481 CACCCCCCCCCCNCCCCCCCCCNACCCCCCCCCCCCCCCCCCNCCCCCCCCC 422

D_b

QY 331 AGCTGCGCCTCGCCGCGCGCCCAACCCAGGCCACCGAGAATGCACCTATG 390
421 ACCCCCCCCCCCCCCCCCCCCCCCCCCCNACNCCCAACCCCAACCCCCCCC 362

QY 391 TCACCGAGTCCACCGCGCGC 410

D_b 361 CCCCCCCCCCCCCCCCCCCCCC 342

RESULT 12

```

US-10-424-599-2320
; Sequence 2320, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 2320
; LENGTH: 580
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(580)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_10209C.1
US-10-424-599-2320

```

Query Match	18.1%;	Score 94.6;	DB 12;	Length 580;
Best Local Similarity	54.9%;	Pred. No. 5e-14;		
Matches 169; Conservative	0;	Mismatches 139;	Indels 0;	Gaps 0;

QY	64	CTCCACCGCGCGGCGCCACCATGATTCTCACCACCGTTACGCCACCTCTCGACGACC	123
Db	267	CC	326
QY	124	CCTTCTCCCCCTTCCACACCGCGTCTCTCTCTCTGCCCCCTTCTCTCCACACCGCGG	183
Db	327	CCCCCCCCNNNCTCCCCCCCCCCCCCCCCCCCCCCCCCTCCCCCCNCCCCCCCCCCCC	386
QY	184	TTTCCTCCACCTGCGCATTTCTTGCCCTCGACTCCCCCTTGCTGCTGACCTTTCCATC	243
Db	387	CCCCCCCCCCCCCGNCCCTCCCGCCCCCCCCCCCCCCCCCTCCCCCCCCCCCCCCCC	446
QY	244	TCCACCCCTTCTCTCCCCACGCGCAGCACTAGCTCCCTCTGACCCCTTCTCTCCACA	303
Db	447	NNCCCCCCCCCCCCCCCCCCCCCCCCCTCTCTTCCCCCCCCCCCCCNCAGCCCNCG	506
QY	304	CCCTCACCACCGCGGTCTCCCACTCGAGCTGCCCCCTGCGCGCGCGGCCACACCCCA	363
Db	507	ACCCCCCCCCCCCCCCCCCCCCCGCNCCTCCCCCCCCCCCCCNCCTCCCCCCCC	566
QY	364	GGCCCAAC 371	
Db	567	CCCGNCCC 574	

RESULT 13

```

US-10-424-599-54839
; Sequence 54839, Application US/10424599
; Publication NO. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovacic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424, 599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 54839
; LENGTH: 778
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(778)
; OTHER INFORMATION: unsure at all n locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_20531C.1
US-10-424-599-54839

```

Query Match	18.0%;	Score 94;	DB 12;	Length 778;
Best Local Similarity	52.6%;	Pred. No. 6.9e-14;		
Matches 202;	Conservative 0;	Mismatches 182;	Indels 0;	Gaps 0;

QY 24 CTCCTTACACTCTCCCACTCCAGCGGAGAGCTGAGCAGAGAGCTCCACCGCGCCGGGCCACC 83
 Db 206 CC 265
 QY 84 ATGAGTTCTACCAACCGTTAGCGCCCACTCTCGAGCAGACCCCTTCTTCCCTTCCACCG 143
 Db 266 CCCCCNCCC 325
 QY 144 CCGTCTCTCTCTCTGCCCCCTTCTCTCTCCCAACCGCGGTTCTTCCACCTGCCATTTC 203
 Db 326 CCCCCCCCCCCCCCTCCCCCCCCCCCCCCCCCCCCCCCCCAACCAACCCCCCCCCCCC 385
 QY 204 TTGCCCCTGACTCCCCCCTTGGCTGTGACCCCTTTCATCTCACCCTTCTCTCCCAAG 263
 Db 386 CCCTCTCCCCCCCCCCCCCTTCCACAACAACCTCCCCCCCCCCCCCCCCCTTCCCC 445
 QY 264 CCGAACCACTAGCTCCCTCTCTCGACCCCTTCTCTCTCCACACCCTCACCAGCGGCTCTCC 323
 Db 446 CCCCCCCCCCTGCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCTTCCCAACCAACCCCTTCCC 505
 QY 324 CAACTGAGCTCGCCCTCGCGCGCCCCGCGCCCCCAACACCCCAAGGCCCAACCAAGGAAGTGC 383
 Db 506 CGCCGCCCATCCCCCCCCCAACACCCCAACCCCCCCCCCCCCCCCCCCCCCCCCCC 565
 QY 384 ACCTATGTCAACCGAGTCCACCGGC 407
 Db 566 CCCCCCATTACTTTTCCCACTGTC 589

RESULT 14

```

US-09-804-682-33
; Sequence 33, Application US/09804682
; Patent No. US20020106765A1
; GENERAL INFORMATION:
; APPLICANT: Kinders, Robert
; APPLICANT: Corey, Michael J.
; TITLE OF INVENTION: PAL-18 POLYPEPTIDES, NUCLEIC ACIDS
; TITLE OF INVENTION: ENCODING THE SAME AND METHODS FOR SCREENING FOR OR
; TITLE OF INVENTION: MODULATING THE SAME
; FILE REFERENCE: 130001.406
; CURRENT APPLICATION NUMBER: US/09/804,682
; CURRENT FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 174

```

NUMBER ;

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 1065
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14, 21, 27, 33, 36, 42, 72, 101, 103, 207, 208, 221, 223,
; LOCATION: 227, 235, 240, 242, 243, 247, 248, 259, 263, 269, 273, 278,
; LOCATION: 296, 321, 322, 324, 330, 332, 335, 336, 340, 367, 371, 385,
; LOCATION: 390, 393, 399, 401, 407, 415, 421, 447, 454, 471, 475
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 479, 494, 501, 508, 511, 513, 525, 539, 540, 579, 596, 605,
; LOCATION: 623, 624, 633, 653, 654, 656, 658, 664, 670, 679, 711, 713,
; LOCATION: 721, 729, 735, 743, 747, 748, 754, 776, 778, 779, 780, 783,
; LOCATION: 798, 802, 808, 810, 813, 814, 820, 822, 824, 825, 838
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 842, 847, 888, 900, 909, 910, 916, 926, 927, 943, 948, 962,
; LOCATION: 973, 1002, 1005, 1028, 1029, 1034, 1057, 1065
; OTHER INFORMATION: n = A,T,C or G
US-09-804-682-33

```

Query Match	17.9%;	Score 93.6;	DB 9;	Length 1065;
Best Local Similarity	48.4%;	Pred. No. 8.5e-14;		
Matches 186;	Conservative	0;	Mismatches 198;	Indels 0;
			Gaps	0;

QY	27	CTTCACACTCCCACTCCAGCGGACGTGAGCAGAGAAGAGTCCACCGCGCGGCCACCATTG	86
Db	308	CTCCCCCCCCCNNTNCCCCCNMTNCNNCCNCCCCCCCACCTCTCCCCCCCCCCCN	367
QY	87	AGTTCTACCAcCGTTACGGCCCACTCTCGAAGAACCCCTTCTTCCCTTCCACCGCG	146
Db	368	CTCNCCCCCTCTCTTTCNCCCCCNCCNCTTNCTCTTNCCTTCTTTCANCCCC	427
QY	147	TCCCTCCTCTCTGCCCCCTTCTCTCCCAACCGCGGTTTCCATCCACTGCACCTATTTTC	206
Db	428	CCCCCCCCCCCCCCCCCTTCTCTCNCCCCCCCCCCTTCCCTTNCNCTCNACCCCC	487
QY	207	GCCCTGACTCCCCCTTGCTGTGAACCTTTCCATGTCAACCCCTTCTCCCCAGCGG	266
Db	488	TTCCCCCNTCCCCCNTTCCCTCNTCNCCCCCCCCCCCCCNCCCCCTCTCTTNCCTCC	547
QY	267	ACCACTAGTCCCTCTCTCGACCCCTTCTCTTCCACACCTCACCGACCGGTCCCAA	326
Db	548	CCCCCCCCCTCTCTTCCCCCCCCCTTCCNCCCCCTTCCCTCTCNCCCCCCCCNCC	607
QY	327	CTGAGCTGCGCCTTGCGCGCGCGCGCCCAACCCCAAGGCCACAGAGAAGTGCACC	386
Db	608	CCCCCCCCCCCCCTCNNCCCCCCCCCAACCCCCCCCCCCCCCTTNNCNCCCTCNCC	667
QY	387	TATGTACCGAGTCCACCGGCGC	410
Db	668	CNCCTTCCANCCTCTCTCCCTC	691

```

RESULT 15
US-09-804-682-29
; Sequence 29, Application US/09804682
; Patent No. US20020106765A1
;
GENERAL INFORMATION:
;
APPLICANT: Kinders, Robert
;
APPLICANT: Corey, Michael J.
;
TITLE OF INVENTION: PAL-18 POLYPEPTIDES, NUCLEIC ACIDS
;
TITLE OF INVENTION: ENCODING THE SAME AND METHODS FOR SCREENING FOR OR
;
TITLE OF INVENTION: MODULATING THE SAME
;
FILE REFERENCE: 130001.406
;
CURRENT APPLICATION NUMBER: US/09/804,682
;
CURRENT FILING DATE: 2001-03-12
;
NUMBER OF SEQ ID NOS: 174
;
SOFTWARE: FastSeq for Windows Version 4.0
;
SEQ ID NO 29

```

```

; LENGTH: 1064
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 5, 6, 16, 21, 24, 25, 33, 39, 72, 110, 209, 214, 231, 232,
; LOCATION: 235, 237, 238, 244, 245, 246, 256, 282, 292, 297, 306, 319,
; LOCATION: 321, 323, 330, 334, 340, 349, 354, 355, 363, 372, 376, 378,
; LOCATION: 397, 405, 432, 437, 454, 455, 457, 458, 459, 468, 470
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 485, 487, 488, 494, 496, 499, 511, 524, 527, 552, 557, 562, 566
; LOCATION: 583, 600, 611, 613, 623, 624, 652, 654, 674, 681, 687, 691
; LOCATION: 694, 701, 713, 716, 720, 721, 725, 731, 734, 735, 739, 743
; LOCATION: 744, 781, 782, 785, 789, 799, 803, 821, 823, 847, 852
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 858, 878, 884, 886, 896, 897, 901, 917, 926, 932, 939, 948
; LOCATION: 957, 961, 965, 981, 991, 993, 1001, 1002, 1005, 1011, 1018,
; LOCATION: 1043, 1047, 1049, 1051, 1054, 1056
; OTHER INFORMATION: n = A,T,C or G
US-09-804-682-29

```

Query Match	17.6%;	Score 92.2;	DB 9;	Length 1064;
Best Local Similarity	49.7%;	Pred. No. 1.9e-13;		
Matches 169;	Conservative 0;	Mismatches 171;	Indels 0;	Gaps 0;

QY		3Z	CAC	TCCCACTCCAGCGGCGACGTGAGCAGAGAGCTCCACC	GCGCGCCGGCCACCATGATTC	91
Db		472	CCCCCCCCCCCC	CNNMCCTGCNCTNCTCTCCCCCCCCCNCCC	CNCNCCC	531
QY		92	TCACA	CGTTAGCGGCCACCTCCTGACGACCCCTTCTTCCCTTCCACCGCGCTCTC		151
Db		532	CCCCCCTCTTA	CCCCCTCCNCCCNCCCNCCCAAACCCCCCCCCCTCNTCTCCCT		591
QY		152	CTCCTCTG	CCCCCTTCTCTTCCCAACCGCGGTTTCTCTCACTGACCATTCTTGCCCT		211
Db		592	CCCTCC	CNTCCCCCCTCNCNCCCCCCCCCNNTCCCTCCGCCCCCCCCCCC		651
QY		212	CGACT	CCCCCTTGGTGTGAACCTTTCCATCTCACACCCTTCTCCCAAGCGACAC		271
Db		652	NCNCCCCCTCC	CTCCCCCCCCCNCCCTCNCNCCCNCTTGNCCCCCNCCCCCTGACC		711
QY		272	TAGT	PCCCTCTGACCCCTTCCCTCCACACCCCTCAACGACCGGCTCTCCCACTGA		331
Db		712	TNCCNCCNNCT	CNCCCTCNCNCCNCCCNCTTCCCCCCCCCTAACCCCCCTTCCCC		771
QY		332	GCTG	GCCTCGCCGCGCGCCGCGCCCAACCCCAAGGCCAC		371
Db		772	TCCCCC	CCCNCCNCCCNCTCCCCCCCCCNCCCNCCCCCCCC	811	

Search completed: April 14, 2004, 14:27:09
Job time : 305 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 14, 2004, 14:05:59 ; Search time 2717 Seconds

(without alignments)
5748.222 Million cell updates/sec

Title: US-09-669-817A-4

Perfect score: 523

Sequence: 1 atctgacagcagatcacca.....gaccgcaagtgtgaagtggaa 523

Scoring table: IDENTITY_NUC

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vr1:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	454.6	86.9	461	9	AU173524 AU173524
2	205.4	39.3	527	10	BF421014 BF421014 FM1_5_B11
3	203.8	39.0	601	14	CD463955 CD463955 ETH1_47_D
4	192.6	36.8	591	14	CA729362 CA729362 wdlc.pk0

5	191.6	36.6	518	14	CA231146	CA211146	SCQSF303
6	191.2	36.6	843	28	CC345900	CC345900	OGPA181TH
7	189.6	36.3	564	13	BU092717	BU092717	946157A06
8	189.6	36.3	604	14	CA831821	CA831821	1117024B0
9	172.8	33.0	509	14	CD429988	CD429988	ETH1_16_D
10	162.2	31.0	728	13	BQ762448	BQ762448	EBR001_5Q
11	160	30.6	483	12	BG356257	BG356257	EM1_23_D0
12	153.6	29.4	658	14	CD213737	CD213737	H81_52_G0
13	152	29.1	662	14	CA168889	CA168889	SCJFSB101
14	150.4	28.8	657	14	CA221736	CA221736	SCRFPL400
15	150.4	28.8	667	14	CA220794	CA220794	SCBGF405
16	148.8	28.5	491	13	BU099577	BU099577	WHE3309_B
17	147.6	28.2	830	29	CC665692	CC665692	OGPA181TV
18	144	27.5	978	14	CA226699	CA226699	SCVPFL304
19	142.6	27.3	622	13	CA091304	CA091304	SCUTAM200
20	141.8	27.1	686	14	CA212362	CA212362	SCSFAD111
21	141.4	27.0	416	13	BU197746	BU197746	946163H05
22	140.2	26.8	655	13	CA067901	CA067901	SCQSFAD105
23	139.6	26.7	718	14	CA209061	CA209061	SCEZAD1C0
24	135.4	25.9	371	28	BZ590427	BZ590427	3590_1_77
25	133.2	25.5	749	14	CA239979	CA239979	SCSBL406
26	131.6	25.2	838	9	AA979751	AA979751	MEST1-D3
27	131.4	25.1	489	14	CA237482	CA237482	SCRFPL504
28	131	25.0	789	14	CA224948	CA224948	SCCCAM2C1
29	130.4	24.9	527	28	BH408147	BH408147	1007056D1
30	130.4	24.9	543	14	CF646081	CF646081	3530_1_10
31	130.4	24.9	591	28	BZ584591	BZ584591	3590_1_58
32	130.4	24.9	1607	11	AY103721	AY103721	Zea_mays
33	130.2	24.9	790	14	CB653015	CB653015	OSJNEC03K
34	128.8	24.6	496	12	BM099211	BM099211	EBES01_SQ
35	123.4	23.6	489	12	BJ303004	BJ303004	BI098190
36	122	23.3	447	13	BU998282	BU998282	HI10J01r
37	117.6	22.5	413	12	BI098190	BI098190	IP1_29_C0
38	117.6	22.5	461	12	BI098181	BI098181	IP1_29_B0
39	116.4	22.3	609	13	CA083273	CA083273	SCEPAM201
40	114.8	22.0	530	14	CD568421	CD568421	3529_1_11
41	113.6	21.7	697	14	CD670649	CD670649	OSMR064_5
42	112	21.4	844	14	CB907084	CB907084	tric081xe
43	111.2	21.3	843	28	AZ200896	AZ200896	SP_1011_A
44	111	21.2	992	29	CNS04R5Q	AL303479	Tetraodon
45	110.2	21.1	1292	13	BQ070431	BQ070431	AGENCOURT

ALIGNMENTS

RESULT 1
LOCUS AU173524 461 bp mRNA linear EST 03-APR-2002
DEFINITION AU173524 Rice root Oryza sativa (japonica cultivar-group) cDNA
clone R3620, mRNA sequence.
ACCESSION AU173524
VERSION AU173524
KEYWORDS AU173524.1 GI:12623311
SOURCE EST.
ORGANISM Oryza sativa (japonica cultivar-group)
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE 1 (bases 1 to 461)
AUTHORS Sasaki,T. and Yamamoto,K.
TITLE Rice cDNA from root (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program, Kannondai 2-1-2, Tsukuba, Ibaraki
305-8602, Japan
Tel: 81-298-38-7441
Fax: 81-298-38-7468
Email: tsasaki@abr.affrc.go.jp, URL: http://rgp.dna.affrc.go.jp/
PROJECT = 'RGP'.
location/Qualifiers
1..461

FEATURES
source

/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultiivar="Niponbare"
/db_xref="taxon:39947"
/clone="R3620"
/clone_1ib="Rice root"
/note="Prepared from seedling root."

ORIGIN

Query Match 86.9%; Score 454.6; DB 9; Length 461;
Best Local Similarity 99.1%; Pred. No. 8.2e-55;
Matches 457; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 10 GCAGATCAACCAAGTCTCTTCACTCCCACTCCAGCGGAGCTGAGCAGAGCTCCAC 69
DB 1 GCAGATCAACCAAGTCTCTTCACTCCCACTCCAGCGGAGCTGAGCAGAGCTCCAC 60
QY 70 CGCGCGCGCGCCACCAAGTCTCTCAACCCGTTACGCCCACTCTCGACGACCCCTTCT 129
DB 61 CGCGCGCGCGCCACCAAGTCTCTCAACCCGTTACGCCCACTCTCGACGACCCCTTCT 120
QY 130 TCCCTTCCCAACCGCGCT 189
DB 121 TCCCTTCCCAACCGCGCT 180
QY 190 CCACTGCGCATTTCTTGGCCCTGACTCCCGCTTGGCTGAGCCCTTCCATCTCCACC 249
DB 181 CCACTGCGCATTTCTTGGCCCTGACTCCCGCTTGGCTGAGCCCTTCCATCTCCACC 240
QY 250 CTTCTCTCCCAACCGCGCT 309
DB 241 CTTCTCTCCCAACCGCGCT 300
QY 310 CCGACCGCGCTCTCCCACTCGAGCTCGCCCTCGCGCGCGCGCGCGCGCGCGCGCGCA 369
DB 301 CCGACCGCGCTCTCCCACTCGAGCTCGCCCTCGCGCGCGCGCGCGCGCGCGCGCGCA 360
QY 370 CCAAGCAGGAAGTGCACCTATGTCACCGAGTCCACCGCGCGCATGTGCAAGTGAACCA 429
DB 361 CCAAGCAGGAAGTGCACCTATGTCACCGAGTCCACCGCGCGCATGTGCAAGTGAACCA 420
QY 430 AGGACAAAGCTTCGCGCGCGCGGAGGGTGTCTCAAGTGGAG 470
DB 421 AGGACAAAGCTTCGCGCGCGCGGAGGGTGTCTCAAGTGGAG 461

RESULT 2
BF421014 527 bp mRNA linear EST 28-NOV-2000
LOCUS FM1.5.B11.b1.A003 Floral-induced Meristem 1 (FM1) Sorghum

DEFINITION propinquum cDNA, mRNA sequence.
ACCESSION BF421014
VERSION BF421014.1 GI:11409003

KEYWORDS EST.
SOURCE Sorghum propinquum
ORGANISM Sorghum propinquum

REFERENCE 1 (bases 1 to 527)
AUTHORS Cordonnier-Pratt, M.-M., Gingle, A., Sudman, M., Marsala, C. and Pratt, L.H.
TITLE An EST database from Sorghum: floral-induced meristems
JOURNAL Unpublished (2000)

COMMENT Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

Sequences have been trimmed to exclude PolyA, vector and regions
below Phred quality 16. The threshold for highest quality sequence

is 20.
Seq primer: JEN REV
High quality sequence stop: 480
POLYA=No.

FEATURES

SOURCE

Source

location/Qualifiers
1..527
/organism="Sorghum propinquum"
/mol_type="mRNA"
/db_xref="taxon:132711"
/clone_1ib="Floral-induced Meristem 1 (FM1)"
/note="Organ: Floral-induced meristems; Vector:
Bluescript II from Lambda Zap II; Site_1: XhoI; Site_2:
EcoRI; mature plants were placed in a growth chamber for
15 days with 16 hr darkness and 8 hr light (flowering is
induced by short-day conditions); 16 days after being
returned to the greenhouse under natural long days during
late April/early May, meristems were harvested The
library was made from poly-A RNA in the cloning vector
lambda Zap II. Clones to be sequenced were prepared by
mass excision."

ORIGIN

Query Match 39.3%; Score 205.4; DB 10; Length 527;
Best Local Similarity 70.4%; Pred. No. 1.1e-19;
Matches 323; Conservative 0; Mismatches 121; Indels 15; Gaps 3;

QY 80 CACCATGAGTCTTCAACCAAGCTTACGCCCACTCTCTGACGACCCCTTCTCCCTTCCC 139
DB 50 CGCATGAGCTTCCCAACCGCACTTCTCTCCGCTCTCTGACGACCCCTTCTTACCTTCCC 109
QY 140 ACCGCGG-----TCTCTCTCTCTCTCTGCGCCCTTCTCTCTCTCTCTCTCTCTCT 193
DB 110 ACCGCGGCACTTCT 169
QY 194 CTGCCATTTCTTGGCCCTGACTCCCTT-----CGCTGCTGACCCCTTTCATCTCCA 247
DB 170 GTCCCGTTCCCGCAGCTGAGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGA 229
QY 248 CCCCTTCTCTCCCAACCGCGCACTAGCTCCCTCTCTGACCCCTTCTCTCTCTCTCTCA 307
DB 230 CCCCTTCT 289
QY 308 CACCGACCGCTCTCCCACTGAGCTCGCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGAG 365
DB 290 CACCGACCGCTCTCGCGCGCTGAGCTCGCCGTGCGCAACCGCGCGCGCGCGCGAGCC 349
QY 366 -CCACAGCAGGAAGTGCACCTATGTCACCGAGTCCACCGCGCGCATGTGATGAGAC 424
DB 350 CCGCGCAGCGGAGAGCGACCTAGTCAACGAGCGCGCGCGCGCGCGCGCGCGAGTGA 409
QY 425 CACCGAGCAGAGCTCTCGCGCGCGGAGAGAGGTGCTCAAGTGGAGGCTGAGCTGACTC 484
DB 410 CTCTGTGAGAAGCGCGCGCGCGGAGAGCGAGCTCAAGTGGAGGCGCGAGATCAAGTC 469
QY 485 CCCCTACGAGATGATTCGACCGCAAGTGAAGTGGGA 523
DB 470 CCGCGCAGCAGAGGCTTCAACCGCAAGTGAAGTGGGA 508

RESULT 3
CD463955

LOCUS

DEFINITION

ETH1_47_D07.g1.A002 Ethylene-treated seedlings Sorghum bicolor cDNA

clone ETH1_47_D07_A002 5', mRNA sequence.

ACCESSION CD463955

VERSION CD463955.1 GI:31385223

KEYWORDS EST.

SOURCE Sorghum bicolor (sorghum)

ORGANISM Sorghum bicolor

REFERENCE 1 (bases 1 to 601)

CD463955 601 bp mRNA linear EST 04-JUN-2003

ETH1_47_D07.g1.A002 Ethylene-treated seedlings Sorghum bicolor cDNA

clone ETH1_47_D07_A002 5', mRNA sequence.

CD463955

CD463955.1 GI:31385223

EST.

Sorghum bicolor (sorghum)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Sorghum.

AUTHORS

Cordonnier-Pratt, M.-M., Wentzel, V., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C., Sun, F., Sullivan, R., Shah, M., Salzman, R., Chua Tan, N., Gonzalez, M., Lane, S., Miller, V., Nanda, P., Olaseinde, O., Eastman, A. and Pratt, L.H.

TITLE

An EST database from Sorghum: 1-aminocyclopropane-1-carboxylic acid (ACC)-treated seedlings

JOURNAL

Unpublished (2003)

COMMENT

Other ESTs: ETH1_47_D07.b1_A002
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.

Seq primer: Sugs (CTTGTGCTCTAAAGCTGCG).

FEATURES

source

Location/Qualifiers
1..601
/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="BTx623"
/db_xref="taxon:4558"
/clone="ETH1_47_D07_A002"
/lab_host="DH10B-T1 phage-resistant E. coli"
/note="Vector: pME185-FL3; Site_1: XhoI; Site_2: XhoI; The library was prepared from polyA+ RNA from seedlings grown in hydroponic culture. At 8 days of age, medium was supplemented with 0.5 mM 1-aminocyclopropane-1-carboxylic acid (ACC) to induce endogenous ethylene (ETH) production. Roots and shoots were harvested after 27 and 72 hr and material from both time points was combined prior to RNA isolation. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME185-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

ORIGIN

Query Match 39.0%; Score 203.8; DB 14; Length 601;

Best Local Similarity 70.2%; Pred. No. 1.9e-19;

Matches 322; Conservative 0; Mismatches 122; Indels 15; Gaps 3;

```
QY 80 CACCATGAGTTCTCACACCGTTAGCCCACTCTCTGACGAGACCCCTTCTCCCTTCCC 139
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 88 CGCCATGGCTTCCACCGCCACCTCTCCGCTCCTCTCGACAACCCCTTCTTCCCTTCCC 147

QY 140 ACCGCG-----TCCTCTCTCTCTGCCCCCTTCTCTTCCCAACCGCGTTTCTTCAC 193
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 148 ACCGCGCCACCTTCTCTCATCTCTGCCCCGTTCTCTTGAACGGGGTTAGACCAACC 207

QY 194 CTGCCATTTCTTGGCCCTGACTCCCTT-----CGTGTGACCCCTTTTCATCTCCA 247
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 208 GTCCCGCTTCCCGACCTGACGACTCTCTCTGCTCGGCTCGGCTTCGACCGCGCTCGA 267

QY 248 CCCCCTTCTCCCAACGCGGACCACTAGCTCCCTCTCTGACCCCTTCTCTCTCAACCT 307
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 268 CCCCCTTCTTGGCTTCCCGCCACCCCGTCCCTTACACTCGCTCTCTCTCGGACCT 327

QY 308 CACCGACGCGTCTCCCACTGAGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGAG-- 365
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 328 CACCGACGCGTCTCCCGCTGAGCTGCGCTGCGCGCGCGCGCGCGCGCGCGAGCC 387

QY 366 -CCACGACGAGAGTGCACCTATGTACCGAGTCCACCGGCGCGCATGTCAAGTGAC 424
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Db

388 CCCCCCACCCGGAGCGGACCTTACGTACCGACGCCGCGCGCGCGCAAGTCAAGTGAC 447

QY

425 CACCGAGACAGCCTCTGCGCGCGCGGAGAGGGTGCTCAAGTGGGAGGCTGAGCTGACTC 484

Db

448 CTGTGTGAGAACCGCGCGCGCGGAGCGGACGCTCAGGTGGAGGCGGAGATCAAGTC 507

QY

485 CCCCTACGAGATGATTGACCGCGAGTGGAAGTGGA 523

Db

508 CCCCCGACGACGCGCTTCGACCGCAAGTGGAAGTGGA 546

RESULT 4

LOCUS

CA729362 591 bp mRNA linear EST 26-NOV-2002

DEFINITION

wdlc.pk006.d7 wdlc Triticum aestivum cDNA clone wdlc.pk006.d7 5'

ACCESSION

CA729362 GI:25451365

VERSION

CA729362.1

KEYWORDS

EST.

SOURCE

Triticum aestivum (bread wheat)

ORGANISM

Triticum aestivum

REFERENCE

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

AUTHORS

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

TITLE

Poideae; Triticaceae; Triticum.

JOURNAL

1 (bases 1 to 591)

COMMENT

Tingey, S.V., Powell, W., Wolters, P., Dolan, M., Hainey, C., Yuan, Z.,

REFERENCE

Miao, G., Caraher, N. and Hanafey, M.K.

JOURNAL

Dupont Wheat cDNA Sequence

COMMENT

Unpublished (2002)

AUTHORS

Contact: Scott V. Tingey

TITLE

Crop Genetics

JOURNAL

E. I. Dupont de Nemours and Company

COMMENT

1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA

REFERENCE

Tel: 302-631-2602

AUTHORS

Fax: 302-631-2607

TITLE

Email: Scott.V.Tingey@USA.dupont.com

COMMENT

Seq primer: M13.

FEATURES

source

Location/Qualifiers
1..591
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wdlc.pk006.d7"
/tissue_type="inflorescence"
/lab_host="DH10B"
/clone_lib="wdlc"
/note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2: XhoI; Wheat (Triticum aestivum, H1 line) developing inflorescence +/- 4 cm"

ORIGIN

Query Match 36.8%; Score 192.6; DB 14; Length 591;

Best Local Similarity 64.5%; Pred. No. 7.1e-18;

Matches 320; Conservative 0; Mismatches 170; Indels 6; Gaps 2;

```
QY 31 ACACTCCCACTCCAGCGGAGTGAGCAGAGAGCTCCACCGCGCGCGGACCATGAGTT 90
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 57 ACACTCCCTCCCACTCCCGTCCGACACCGTCCGCTCCGCGCGCGCGGCGCATGGCT 116

QY 91 CTCACCAACGTTAC--GCCACCTCTCTGACGACGACCCCTTCTTCCCTTCCACCGCGCT 147
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 117 CTTCCACGCGCACCTGATGCGCTGCTGACGACGACCCCTTCTTCCCTTCCCGGCCAC 176

QY 148 CCTCTCTCTGCCCCCTTCTCTTCCCAACCGCGGTTTCTTCCCACTGCGCATTTCTTG 207
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 177 CACCACTCTCTCTCTCTCTCACTTCTGCCCCCTTCTTCCCAACCCCATCC 236

QY 208 CCCTGACTCCCTTCTGCTGCTGCTGCTTTCATCTTCACTTCCATCCCTTCTTCCCAAGCGGA 267
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db 237 CCTTGGCTTCCCGCACACCGCGGACCTCTGACCTTACCTTCCCGGACCCCTTCT 296

QY 268 CCACTAGTCCCTCTGACCCCTTCTCTCCACACCTTCAACCGGCGGTCTCCGAC 327
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
```

Db 297 TCGCCCTCCCTCCCGCAGCCCTTCTCTCCACGACCTCAGGACCGCGTGGCGGCC 356
Qy 328 TCGAGCTGCGCTCCG 387
Db 357 TCGAGCTGCG 413
Qy 388 ATGTACCGAGTCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 447
Db 414 CGGCCACGAGCG 473
Qy 448 GCGAGAGGCTCTCAAGTGGAGGCTGAGCTGAGTCCCGCTACGACGATGATTCGACC 507
Db 474 GGGCGCGCGCTCAAGTGGAGGCTGAGTCCCGCTACGACGATGATTCGACC 533
Qy 508 GCAAGTGAAGTGGGA 523
Db 534 GCAAGTGGCGGTGGGA 549

RESULT 5
CA231146

LOCUS CA231146 518 bp mRNA linear EST 25-SEP-2003
DEFINITION SCQSL3031G02.g Saccharum officinarum FL3 Saccharum officinarum
CDNA clone SCQSL3031G02 5', mRNA sequence.

ACCESSION CA231146 GI:35295392
VERSION CA231146.1 GI:35295392
KEYWORDS EST.

SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.
1 (bases 1 to 518)

AUTHORS Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE The libraries that made SUCEST
JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

COMMENT Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089

Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.fcav.unesp.br
plate: 031 row: G column: 02
Seq primer: T7 Promoter Primer.

FEATURES
source Location/Qualifiers
1..518
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCQSL3031G02"
/lab_host="DH10B"
/clone_lib="Saccharum officinarum FL3"
/note="Organ: Base of developing inflorescence (5cm-long);
vector: pSport1; Site_1: SalI; Site_2: NotI; An
unidirectional cDNA library generated from (Base of
developing inflorescence (5cm-long)). cDNA was prepared
from polyA+ mRNA using Superscript Plasmid System Kit
(Invitrogen). The double-strand cDNAs were fractionated
in a sepharose CL-2B 40cm-columns and fragments sizing
between 0.8 and 1.5 Kb were directionally cloned into the
vector. Details of each source of RNA and library
construction can be obtained at
http://sucest.lad.ic.unicamp.br/public"

ORIGIN

Query Match 36.6%; Score 191.6; DB 14; Length 518;
Best Local Similarity 68.8%; Pred. No. 9.6e-18;
Matches 317; Conservative 0; Mismatches 120; Indels 24; Gaps 3;

Qy 80 CACCATGAGTTCTCACCACCGTTAGCGCCACCTCTCTGAGAGACCCCTTCTTCCCTTCCC 139
Db 58 CGCATGGCTTCCACCGCCACTTCTCTCGCTCTGACGACCCCTTCTTCCCTTCCC 117
Qy 140 ACCGCGC---TCTCTCTCTCTCTGCGCTTCTT-----CTCCGAC 178
Db 118 ACCGCGCGACCTTCT 177
Qy 179 CGCGTTTCTCTCACCCTGCGCATCTTCTGCGCTCTGAGTCTCTCTCTCTCTCTCTCT 238
Db 178 CCGGTCTCTCTCTCACCCTGCGCATCTTCTGCGCTCTGAGTCTCTCTCTCTCTCTCTCT 237
Qy 239 CCATCTCACCCT 298
Db 238 GCGCTCAGCCCTTCT 297
Qy 299 CCACACCTCAGCCGCGCT 358
Db 298 CCGGACCTCAGCCGCGCT 357
Qy 359 CCCGAG---CCACACGAGAGTGTACCTATGTACAGGATCCACCGCGCGCATGT 415
Db 358 CCGGAGCG 417
Qy 416 CAAGTGACACCGAGGACAAAGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 475
Db 418 CAAGTGACCT 477
Qy 476 GCTGACTCCCT 516
Db 478 GATCAAGTCCCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 518

RESULT 6
CC345900

LOCUS CC345900 843 bp DNA linear GSS 16-MAY-2003
DEFINITION OGP181TH ZM_0.7_1.5_KB Zea mays genomic clone ZMMBma0350M17,
genomic survey sequence.

ACCESSION CC345900 GI:30815292
VERSION CC345900.1 GI:30815292
KEYWORDS GSS.

SOURCE Zea mays
ORGANISM Zea mays

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 843)

AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nuneberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)

TITLE Contact: Cathy Whitelaw
JOURNAL
COMMENT TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TR
Class: sheared ends.

FEATURES
source Location/Qualifiers
1..843
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBma0350M17"
/clone="ZM_0.7_1.5_KB"
/note="Vector: pBSK-; Site_1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN

Query Match 36.6%; Score 191.2; DB 28; Length 843;

Best Local Similarity 67.1%; Pred. No. 1.2e-17;
Matches 306; Conservative 0; Mismatches 138; Indels 12; Gaps 2;

OY 80 CACCATGAGTTCACCAACCGTTACGCCCAACCTCCTCGACGACCCCTTCTTCCCTTCCC 139
 DB 222 CGCCATGGCTTCCCATCGCCACTTCTACGCTCCTCGAAGACCCCTTCTCCCCCTTCCC 281

QY 140 ACCGCGTCTCTCTCTCTCTGCCCCCTTCTCTCTCCCAACCCGCGCTTCTCTCAACTGCC 199
 Db 282 ACCGCGGCCAACCTTCTCTCTCTCTCTCTGCCCCGTTCTCTCTGAACGCGCTCTCAAGTTACCC 341
 QY 200 ATTCTTGGCCCTCGACTCCCCCTTGGCTGCTGAACCTTTCCATCTCAACCCCTTCTCC 259
 Db 342 GTCCCCGTTCCCCGCGCCTCCCCCATTTCCCGAAGCTAGATGATCTCTCTGCCCCCGC 401
 QY 260 CACGCCGACCACTAGTCTCCCT-----CCTGACCCCTTCTCTCTCAACCCCTCAC 310
 Db 402 GCGGCGCTCGACCCCTTCTTGGCTTCAACCCCAACCCCGCTCATTTCTCGAGACCTCAC 461
 QY 311 CGACCGGCTTCCCACTCGAGCTGCGCCTGCGCGCCGCGCCCAACCCAG--GCC 367
 Db 462 CAACCGTGTGCGCGCACTCGAGGTGCGCGCTGCGCAACGCGCGCGCGCCCGGACCCCC 521
 QY 368 CACCAGCAGGAAGTGCACCTATGTCAACGAGTCCACCGGCGCATGGTCAAGTGAACAC 427
 Db 522 CGCCACCCGGAAGCGCACCTATGTCACTGAGGCCGCGCGCCGCAAGTCAAGTGAACTTC 581
 QY 428 CGAGGACAAGCCTCGCGCCGCGCAGAGGCTGCTCAAGTGGAGGCTGAGCTCGACTCCCC 487
 Db 582 CGTCGAGAAGCGCGCGCGCGGAGCGGACGCTCAAGTGGAGGCGCGGATCAAGTCCCC 641
 QY 488 CTACGACGATGGATTGCAACCGCAAGTGAAGTGGGA 523
 Db 642 CGACGACGACGCGCTTGAACCGCAAGTGAAGTGGGA 677

RESULT 7
 BU092717
 LOCUS
 DEFINITION BU092717 564 bp mRNA linear EST 29-AUG-2002
 946157A06.y1 946 - tassell primumdium prepared by Schmidt lab Zea
 mays cDNA, mRNA sequence.
 ACCESSION BU092717
 VERSION BU092717.1 GI:22542279
 KEYWORDS EST.
 SOURCE Zea mays
 ORGANISM Zea mays
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 564)
 Walbot, V.
 Maize ESTs from various cDNA libraries sequenced at Stanford
 University
 Unpublished (1999)
 JOURNAL
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 946157 row: A column: 06.
 FEATURES
 Location/Qualifiers
 1..564
 Source

```

/organism="Zea mays"
/mol_type="mRNA"
/cultivar="OH43"
/db_xref="taxon:4577"
/tissue_type="tassels"
/dev_stage="just after the transition from vegetative to
inflorescence development"
/lab_host="XLOLR"
/clone_lib="946 - tassel primordium prepared by Schmidt

```

```
lab"
/site="Organ: tassels, Vector: HybridZAP, site 1: EcoRI;
/site 2: XhoI; George Chuck dissected immature tassels
between 1mm and 3mm. Sharon Stanfield prepared the cDNA
library in HybridZAP. Sample insert size range was 350 bp
to 3 Kb with a 1 Kb average."
```

Query Match	36.3%;	Score 189.6;	DB 13;	Length 564;
Best Local Similarity	66.9%;	Pred. No. 1.9e-17;		
Matches 305; Conservative	0;	Mismatches 139;	Indels 12;	Gaps 2;

QY	80	CACCATGAGTTCACACCACCGTTACGCCCACTCCTCGACGACCCCTTCTTCCCCCTGCC	139
Dp	17	CGCCATGGCTTCCCATCGCCAATTCCTAAGCCCTCCTCGAAGACCCCTTCTCCCCCTTCCC	76
QY	140	ACGCGCGTCTCTCTCTCTCTGCCCCCTTCTCTTCCCCAACCCCGGTTCCTCCACTTGCCC	199
Dp	77	ACGCGCGCACCTTCTCTCTCTCTCTGCCCCGTTCTCTTGACGGCGTCTCAGATTACCC	136
QY	200	ATTCTTGCCCTCGACTCCCCCTTGCTGCTGACCTTTGCATCTCCACCCCTTCTCTCC	259
Dp	137	GTCCTCGTTCCCGCGCGCTCCCAATTCCTCCGACCTAGATGATCTCTCTGCCCCCGCG	196
QY	260	CACGCCGACCACTAGCTCCCT-----CTCGACCCCTTCTCTCTCCACACCCTCAC	310
Dp	197	GCCGGCGCTGACCCCTTCTTGCCCTTCAACCCCAACCCGCTCATTTCTCCGAGACTCAC	256
QY	311	CGACCGGTTCTCCCACTCGAGCTGCCCCCTGCGCGCGCCCAACCCCAAG---GCC	367
Dp	257	CAACCGTGTCCCGCACTCGAGTGTGCGCTCGCCACCGCGCGCGCCCGGACCCCC	316
QY	368	CACCAGCAGAATGCACTTATGTACCGAGTCCACCGGCGCATGTGTCAGTGAACAC	427
Dp	317	CGCCACCCGGAAGCGCACTATGTCACTGAGCGCGCGCGGCGCAAGTCAAGTGACTTC	376
QY	428	CGAGACAAGCTCTCGCGCGCGCGGAGAGGCTGCTCAAAGTGGAGGCTGAGCTGCATCCCC	487
Dp	377	CGTCGAGAAACCGCGCGCGGGGAGCGGACGCTCAAGTGGAGGCGGAGATCAAGTCCCC	436
QY	488	CTACGACGATGATTCGACCCGCAAGTGGAGTGGGA	523
Dp	437	CGACGACGACGGCTTCGACCGCAAGTGGAGTGGGA	472

RESULT 8
LOCUS CA831821
DEFINITION 1117024B07.y1 1117 - Unigene V from Maize Genome Project Zea mays
604 bp mRNA linear EST 12-DEC-2002
CDNA, mRNA sequence.
ACCESSION CA831821
VERSION CA831821.1 GI:26559586
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 604)
REFERENCE Walbot, V.
AUTHORS Maize ESTs from various cDNA libraries sequenced at Stanford
TITLE University
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V

FEATURES	Location/Qualifiers
source	1. .604

ORIGIN

/organism="Zea mays"
/mol_type="mRNA"
/db_xref="taxon:4577"
/clone_lib="1117 - Unigene V from Maize Genome Project"
/note="This library represents the unique genes found in the fifth round of EST sequencing at Stanford University for the maize genome project. Sequences are present from library 946. Contigs were assembled using ZmBAssembler and 2 representatives from each contig were selected for the Unigene set. All singlets were also selected."

Query Match 36.3%; Score 189.6; DB 14; Length 604;
Best Local Similarity 66.9%; Pred. No. 1.9e-17;
Matches 305; Conservative 0; Mismatches 139; Indels 12; Gaps 2;

QY 80 CACCATGAGTCTCTACACCGGTTACGCCCACTCTCTCGACGACCCCTTCTCCCTTCCC 139
DB 61 CGCCATGGCTTCCCATCGCCCACTTCTTAAGCCTCTCGAAGACCCCTTCTCCCTTCCC 120
QY 140 ACCGCGCT 199
DB 121 ACCGCGCGACCTTCT 180
QY 200 ATTCTTGGCTCTGACTCCCT 259
DB 181 GTCCCGCTTCT 240
QY 260 CACGCGCGACCTAGTCT 310
DB 241 GCGCGCGCTGACCCCTTCT 300
QY 311 CGACCGCGCT 367
DB 301 CAACGCTGCTCGCCT 360
QY 368 CACGCGCGAGTGTGACCTATGTACCGAGTCCACCGCGCGCATGTCAAGTGACAC 427
DB 361 CGCCACCGCGAGCGACCTATGTACCGAGTCCACCGCGCGCATGTCAAGTGACATTC 420
QY 428 CGAGGACAGCTCTCGCGCGCGAGAGGCTCTCAAGTGAGGAGCTGAGCTGACTCTCCC 487
DB 421 CGTCGAGAGCGCGCGCGCGAGGCTCTCAAGTGAGGAGCTGAGCTGACTCTCCC 480
QY 488 CTACGAGATGATTCGACCGCAAGTGAGTGGGA 523
DB 481 CGACGAGAGCGCTTCGACCGCAAGTGAGTGGGA 516

RESULT 9
CD429988 509 bp mRNA linear EST 03-JUN-2003
LOCUS ETH1_16 D10.G1 A002 Ethylene-treated seedlings Sorghum bicolor CDNA
DEFINITION clone ETH1_16 D10 A002 5', mRNA sequence.
ACCESSION CD429988
VERSION CD429988.1 GI:31345631
KEYWORDS EST.
SOURCE Sorghum bicolor (sorghum)
ORGANISM Sorghum bicolor
REFERENCE Sorghum bicolor; Streptophyta; Embryophyta; Tracheophyta; Eukaryota; Viridiplantae; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoidae; Andropogoneae; Sorghum.
AUTHORS Cordonnier-Pratt, M.-M., Wentzel, V., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C., Sun, F., Sullivan, R., Shah, M., Salzman, R., Chua Tan, N., Gonzalez, M., Lane, S., Miller, V., Nanda, P., Olaseinde, O., Eastman, A. and Pratt, L.H.
TITLE An EST database from Sorghum: 1-aminocyclopropane-1-carboxylic acid (ACC)-treated seedlings
JOURNAL Unpublished (2003)
COMMENT Other ESTs: ETH1_16 D10.b1 A002
Contact: Cordonnier-Pratt MM
Laboratory for Genomics and Bioinformatics

FEATURES

The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu
Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.
Seq primer: Sug5 (CTTCTGCTCTAAAGCTGCG).
Location/Qualifiers
1..509
/organism="Sorghum bicolor"
/mol_type="mRNA"
/cultivar="BTx623"
/db_xref="taxon:4558"
/clone="ETH1_16 D10 A002"
/lab_host="DH10B-T1 phage-resistant E. coli"
/note="Vector: pME185-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA+ RNA from seedlings grown in hydroponic culture. At 8 days of age, medium was supplemented with 0.5 mM 1-aminocyclopropane-1-carboxylic acid (ACC) to induce endogenous ethylene (ETH) production. Roots and shoots were harvested after 27 and 72 hr and material from both time points was combined prior to RNA isolation. Double-stranded cDNA was cloned. The pME185-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

ORIGIN

Query Match 33.0%; Score 172.8; DB 14; Length 509;
Best Local Similarity 68.0%; Pred. No. 4.3e-15;
Matches 291; Conservative 0; Mismatches 122; Indels 15; Gaps 3;

QY 80 CACCATGAGTCTCTACACCGGTTACGCCCACTCTCTCGACGACCCCTTCTCCCTTCCC 139
DB 82 CGCCATGGCTTCCCAACCGCACTTCTCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 141
QY 140 ACCGCGG-----TCT 193
DB 142 ACCGCGCGCACTTCT 201
QY 194 CTGCCCCATCTTGGCCCTCGACTCCCTCT-----CGCTGTGACCTTTCCATCTCCA 247
DB 202 GTCCCGCTTCCCGCACTTGACGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 261
QY 248 CCCCTTCTCTCCGACCGGACCACTAGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 307
DB 262 CCCCTTCT 321
QY 308 CACCGACCGGCTTCT 365
DB 322 CACCGACCGGCTTCT 381
QY 366 -CCGACGAGAGAGTGTACCTATGTACCGAGTCCACCGGCGCGCATGTCTCAAGTGAG 424
DB 382 CCGCGCGACCGGAGGCGACCTAGCTACCGGCGCGCGCATGTCTCAAGTGAG 441
QY 425 CACCGAGAGAGAGCTTCT 484
DB 442 CTCTGTGAGAGAGAGCGCGCGCGGAGCGGAGCGCTCATGTGGAGGCGCGAGATCAAGTC 501
QY 485 CCCCTAG 492
DB 502 CCGCGAG 509

```
RESULT 10
LOCUS      BQ762448
DEFINITION BQ762448      728 bp      mRNA      linear      EST 26-JUL-2002
EBR001.SQ005_J23_R root, 3 week, hydroponic grown, no treatment,
cv Optic, EBR001 Hordeum vulgare subsp. vulgare cDNA clone
EBR001.SQ005_J23 5', mRNA sequence.

ACCESSION  BQ762448
VERSION     BQ762448.1
KEYWORDS    GI:21970920
SOURCE      Hordeum vulgare subsp. vulgare
ORGANISM    Hordeum vulgare subsp. vulgare
REFERENCE   Hedley,P., Liu,H., Caldwell,D., McCallum,N., Mudie,S., Cardle,L.,
AUTHORS     Ramsay,L., Machray,G., Marshall,D.F.M. and Waugh,R.
TITLE       Development of Barley Transcriptome Resources
JOURNAL     Unpublished (2001)
COMMENT     Contact: Waugh R, Marshall DF
            Genome Dynamics/Computational Biology
            Scottish Crop Research Institute
            Invergowrie, Dundee, DD2 5DA, Scotland, UK
            Tel: 00 44 1382 562731
            Fax: 00 44 1382 562426
            Email: est@scri.sari.ac.uk.
            Location/Qualifiers
            1..728
            /organism="Hordeum vulgare subsp. vulgare"
            /mol_type="mRNA"
            /cultivar="Optic"
            /sub_species="vulgare"
            /db_xref="taxon:112509"
            /clone="EBR001.SQ005_J23"
            /tissue_type="root"
            /dev_stage="3 week"
            /lab_host="DH103"
            /clone_lib="root, 3 week, hydroponic grown, no treatment,
            cv Optic, EBR001"
            /note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;
            Non-normalised library, directionally cloned into pSPORT1.
            Derived from roots of 3 week old hydroponically grown
            unstressed barley plants. Developed as part of the barley
            transcriptome resources of BBSRC/SEERAD funded cereal IGF
            (Investigating Gene Function) project."

ORIGIN
Query Match      31.0%; Score 162.2; DB 13; Length 728;
Best Local Similarity 63.7%; Pred. No. 1.4e-13;
Matches 300; Conservative 0; Mismatches 158; Indels 13; Gaps 3;

QY 66 CCACCGCGCGCGCCACCATGAGTTCTCACCCACCGTTACGCCACCTCTCGACGACCCC 125
    |||||
DB 40 CCGCGCGCGCGCCACCATGAGTTCTCTCCACCGCCACCTCATGCGCTGCTGACGACCCC 99
    |||||

QY 126 TTCTTCCCTTCCACCGCGGTCCTCTCTCTCTGCCCCCTTCTCTCCACCGCGGTT 185
    |||||
DB 100 TTCTTCCCTTCCCGCGCCACCGCTACCTCTCTCTCTCTCTCTGCCCCCTTCTCTC 159
    |||||

QY 186 TCCTCACTGCGCAT-----TCTTGCGCCTGACTGCCCTTGTGCTGACCT 236
    |||||
DB 160 CCACCTCACTCTCATCCCATTCGCTTCCCGACGACGACCTCTGACCTTAACCTC 219
    |||||

QY 237 TTCATCTCAACCCCTTCTCTCCCGACGCGACCACTAGTCTCTCTCTGACCCCTTCTC 296
    |||||
DB 220 CCGCGCGACCCCTTCTCTCTCCACCGCCACCGCCCTTCCCGCCACACCGGACGCC 279
    |||||

QY 297 CTCACACCTTCAACGACCGGCTCTCCCACTGAGCTGCGCTTGGCGCGCGGCCCA 356
    |||||
DB 280 TTCTCTCTCAACGACCTCAACGACCGGCTGCGCGCTGAGCTGCGCGCGCGCGCG 339
    |||||

QY 357 CACCCGAGGCCCAACGACGAGGAGTGAAGTATGTACCGAGTTCAC--CGCGCGCATG 413
    |||||
```

```
RESULT 11
LOCUS      BG356257
DEFINITION BG356257      483 bp      mRNA      linear      EST 06-MAR-2001
sequence.
EM1_23_D03_b1_A002 Embryo 1 (EM1) Sorghum bicolor cDNA, mRNA

ACCESSION  BG356257
VERSION     BG356257.1
KEYWORDS    GI:13238243
SOURCE      EST.
ORGANISM    Sorghum bicolor (sorghum)
REFERENCE   Sorghum bicolor
AUTHORS     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
TITLE       Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
JOURNAL     clade; Panicoideae; Andropogoneae; Sorghum.
            1 (bases 1 to 483)
            Reid,S.P., Cordonnier-Pratt,M.-M., Gingle,A. and Pratt,L.H.
            An EST database from Sorghum: developing embryos
            Unpublished (2000)
            Contact: Cordonnier-Pratt MM
            Laboratory for Genomics and Bioinformatics
            The University of Georgia, Department of Plant Biology
            Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
            Tel: 706 542 1860
            Fax: 706 583 0210
            Email: mmpratt@uga.edu
            Sequences have been trimmed to exclude PolyA, vector and regions
            below Phred quality 16. The threshold for highest quality sequence
            is 20.
            Seq primer: JEN REV
            High quality sequence stop: 475
            POLYA=No.
            Location/Qualifiers
            1..483
            /organism="Sorghum bicolor"
            /mol_type="mRNA"
            /db_xref="taxon:4558"
            /clone_lib="Embryo 1 (EM1)"
            /note="Organ: Embryos germinated for 24 hr; Vector:
            pBluescript II from lambda Zap II; Site 1: XhoI; Site 2:
            EcoRI; The library was made from poly-A RNA in the cloning
            vector lambda Zap II. Clones to be sequenced were
            prepared by mass excision."

ORIGIN
Query Match      30.6%; Score 160; DB 12; Length 483;
Best Local Similarity 67.8%; Pred. No. 2.7e-13;
Matches 274; Conservative 0; Mismatches 115; Indels 15; Gaps 3;

QY 80 CACCATGAGTTCTACCAACCGTACGCCCACTCTCTGACGACCCCTTCTCCCTTCCC 139
    |||||
DB 80 CGCCATGGCTTCCCAACCGCACTTCTCGGCTCTCTGACGACCCCTTCTTCCCTTCCC 139
    |||||

QY 140 ACCGCG-----TCCTCTCTCTGCGCCCTCTCTCTCCGACCGCGGTTCTCTGAC 193
    |||||
DB 140 ACCGCGGCACTTCTCTCATCTCTGCGCCCTCTCTCTCTGACGCGGTTCTGACGACCC 199
    |||||

QY 194 CTGCCATTTCTGCGCTGACTCCCTT-----CGTGTGACCCCTTTCATCTCA 247
    |||||
DB 200 GTCCCGGTTCCCGACCTGACGACCTCTCTCTGCGCTTCCGACCGCGGCTCA 259
    |||||

QY 248 CCGCTTCTCTCCGACCGGCACTAGCTCTCTCTGACCCCTTCTCTCAACCT 307
    |||||
```

Db 260 CCCCTTCTTCGCTTCCCGCCACCCCGTCCCTACACTCCGCTCTCTCCGCGACCT 319

QY 308 CACCGACCGCGCTTCCCACTCGAGCTCGCCCTCGCCGCGCGCCCGCCACACCCGAG-- 365

Db 320 CACCGACCGCGCTCGCGCTCGAGCTCGCGCTCGCCACCGCGCGCGCGCCCGAGCC 379

QY 366 -CCCGACGAGAGAGTGCACCTATGTACCGAGTCCACCGCGCGCGATGTCAAGTGAC 424

Db 380 CCGCGCCACCGGAGAGCGACCTACGTACCGACCGCGCGCGCGAGGTCAAGTGAC 439

QY 425 CACCGAGACAGCCTCGCGCGCGCGAGAGGCTCTCAAGTGG 468

Db 440 CTCTGTGAGAGCGCGCGCGCGGAGCGGACGCTCAGTGG 483

RESULT 12

LOCUS CD213737 658 bp mRNA linear EST 20-MAY-2003

DEFINITION HS1_52_G09.g1 A012 Heat-shocked seedlings Sorghum bicolor cDNA clone HS1_52_G09_A012 5', mRNA sequence.

ACCESSION CD213737

VERSION CD213737.1 GI:30952107

KEYWORDS EST.

SOURCE Sorghum bicolor (sorghum)

ORGANISM Sorghum bicolor

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.

AUTHORS 1 (bases 1 to 658)

REFERENCE Cordonier-Pratt, M.-M., Wentzel, V., Suzuki, Y., Sugano, S., Klein, R.R., Liang, C., Sun, F., Sullivan, R., Shah, M., Buchanan, C.D., Eastman, A., and Pratt, L.H.

TITLE An EST database from Sorghum: heat-shocked seedlings

JOURNAL Unpublished (2003)

COMMENT Other ESTs: HS1_52_G09.b1_A012

CONTACT Cordonier-Pratt MM

LABORATORY Laboratory for Genomics and Bioinformatics

UNIVERSITY The University of Georgia, Department of Plant Biology

PLANT Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA

TEL: 706 542 1860

FAX: 706 583 0210

EMAIL: mmpratt@uga.edu

LIBRARY Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science; plant material and RNA prepared at Texas A & M University; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA.

Seq primer: Sugs (CTTCTGCTCTAAAGCTGCG).

FEATURES

source

1. 658

/organism="Sorghum bicolor"

/mol_type="mRNA"

/cultivar="IS3620C"

/db_xref="taxon:4558"

/clone="HS1_52_G09_A012"

/lab_host="DH10B-TI phage-resistant E. coli"

/clone_lib="Heat-shocked seedlings"

/note="Vector: PME18S-FL3; Site_1: XhoI; Site_2: XhoI; The library was prepared from polyA+ RNA from 6-day-old seedlings grown in hydroponic culture and heat-shocked at 40-42 C for 4 or 24 hr. After heat shock, roots and leaves were harvested and tissues combined for RNA isolation. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the PME18S-FL3 vector (5-prime DraIII site is CACTGTGTG, 3-prime DraIII site is CACCATGTG)."

ORIGIN

Query Match 29.4%; Score 153.6; DB 14; Length 658;

Best Local Similarity 67.4%; Pred. No. 2.3e-12;

Matches 267; Conservative 0; Mismatches 114; Indels 15; Gaps 3;

QY 80 CACCATGAGTTCTTCCACCACCGTTACGCCCACTCTCTCGACGACCCCTTCTCCCTTCCC 139

Db 263 CGCCATGCTTCCACCGCCACTTCTCTCGCTCTCTCGACACCCCTTCTCCCTTCCC 322

QY 140 ACCGCGG-----TCTCTCTCTCTCTGCGCCCTTCTCTCTCCCAACCCCGTTCTCCAC 193

Db 323 ACCGCGCCACCTTCTCTCATCTCTCTGCGCTTCTCTCTGACGGGGTCTTAGACCAACC 382

QY 194 CTGCCCATTTCTGCGCCCTCGACTCCCTCT-----CGCTGTGACCCCTTTCATCTCCA 247

Db 383 GTCCCGGTTCCCGACCTGACGACCTCTCTCTGCTCCCTCGGCTCTCCGACCGGCTCGA 442

QY 248 CCCCTTCTTCCCGACCGCCGACCACTAGCTCTCTCTCTCGACCCCTTCTCTCTCCACACCT 307

Db 443 CCCCTTCTTCCCGCTTCCCGCCGACCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 502

QY 308 CACCGACCGCGCTTCCCACTCGAGCTCGCCCTCGCCGCGCGCCCGCCACACCCGAG-- 365

Db 503 CACCGACCGCGCTCGCGCGCTCGAGCTCGCGCTCGCCACCGCGCGCGCGCCCGAGCC 562

QY 366 -CCCGACGAGAGTGCACCTATGTACAGAGTCCACCGCGCGCGAGTGTCAAGTGAC 424

Db 563 CCGCGCCACCGGAGAGCGACCTACGTACCGACGCGCGCGCGCGCGAGGTCAAGTGAC 622

QY 425 CACCGAGACAGCCTCGCGCGCGCGGAGAGGCTGCT 460

Db 623 CTCTGTGAGAGCGCGCGCGCGGAGGCGACGCT 658

RESULT 13

LOCUS CA168889 662 bp mRNA linear EST 24-SEP-2003

DEFINITION SCJFSB1010D05.g SB1 Saccharum officinarum cDNA clone SCJFSB1010D05 5', mRNA sequence.

ACCESSION CA168889

VERSION CA168889.1 GI:35090764

KEYWORDS EST.

SOURCE Saccharum officinarum

ORGANISM Saccharum officinarum

REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Saccharum.

AUTHORS 1 (bases 1 to 662)

REFERENCE Vettore, A.L., da Silva, F.R., Kemper, E.L. and Arruda, P.

TITLE The libraries that made SUCEST

JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)

COMMENT Contact: Arruda P

CONTACT Centro de Biologia Molecular e Engenharia Genetica

UNIVERSITY Universidade Estadual de Campinas

PLANT Caixa Postal 6010, 13083-970, Campinas SP, Brazil

TEL: 55 19 3788 1137

FAX: 55 19 3788 1089

EMAIL: parruda@unicamp.br

LIBRARY Clone distribution: clone distribution information can be found through the Brazilian Clone Collection Center (BCCC) at <http://www.bcccenter.fcav.unesp.br>

PLATE: 010 row: D column: 05

Seq primer: T7 Promoter Primer.

FEATURES

source

1. 662

/organism="Saccharum officinarum"

/mol_type="mRNA"

/db_xref="taxon:4547"

/clone="SCJFSB1010D05"

/lab_host="DH10B"

/clone_lib="SB1"

/note="Organ: Stalk Bark from adult plants; Vector: pSport1; Site 1: SalI; Site 2: NotI; An unidirectional cDNA library generated from [Stalk Bark from adult plants]. cDNA was prepared from polyA+ mRNA using Superscript Plasmid System Kit (Invitrogen). The

double-strand cDNAs were fractionated in a sepharose CL-2B 40cm-columns and fragments sizing between 0.8 and 1.5 Kb were directionally cloned into the vector. Details of each source of RNA and library construction can be obtained at <http://sucest.lad.ic.unicamp.br/public>

ORIGIN

Query Match 29.1%; Score 152; DB 14; Length 662;
Best Local Similarity 62.6%; Pred. No. 3.8e-12;
Matches 283; Conservative 0; Mismatches 145; Indels 24; Gaps 2;

```
QY 93 CACCACCGTTACGCCCACTCTCGACGACCCCTTCTCCCTTCCACCGCGCTCTCC 152
DB 46 CTCACGCGCTGACGACGACCCCTTCTTCCCATCTTCCCTTCCACGACGACGAC 105
QY 153 TCCTCTGCGCCCTTCTCTCTCCGACCGCGGTTCTCTCCACCTGCGCATTTCTGCGCTC 212
DB 106 TCGTCTGCGCCCTTACATCTCGGCTCTCTCTCTCTCCGACGACGCTTCTCTGAC 165
QY 213 GACTCCCTCTGCTGCTGACCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 272
DB 166 GACGACCCCTTCTTACCCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 225
QY 273 AGTCCCTCTCTGACCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 311
DB 226 ACCTTGACCTGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 285
QY 312 GACCGCGTCTCCCACTGAGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 371
DB 286 GACTTGACCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 345
QY 372 AGCAGGAAGTGACCTATGTACCGAGTCCACCGCGCATGTCAAGTGACCAAGGAG 431
DB 346 CGTCGCAAGTACACTACGCTCCGAGGCGGACGAGGAGGAGTCAAGTGACCGCGGAG 405
QY 432 GACAAGCTCGCGCGCGGAGAGGCTCTCAAGTGAGGCTGAGCTCTCTCTCTCTCT 491
DB 406 GACAGGCGC--CGCGCGGACCGGAGCTCAAGTGAGGAGCGGAGCTCAAGAGCGCTTAC 462
QY 492 GACGATGATTCGACCGCAAGTGAGAGTGGGA 523
DB 463 CACGACGCGCTTCGACCGCAAGTGAGAGTGGGA 494
```

RESULT 14

CA221736 657 bp mRNA linear EST 25-SEP-2003
LOCUS SCRFPL4008G02.g FL4 Saccharum officinarum cDNA clone SCRFPL4008G02
DEFINITION 5', mRNA sequence.
ACCESSION CA221736
VERSION CA221736.1 GI:35276347
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.
1 (bases 1 to 657)
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
The libraries that made SUCEST
Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Email: patruada@unicamp.br
Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
<http://www.bcccenter.fcav.unesp.br>
Plate: 008 row: G column: 02
Seq primer: T7 Promoter Primer.

FEATURES

source

Location/Qualifiers
1. .657
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCRFPL4008G02"
/lab_host="DH108"
/clone_lib="FL4"
/note="Organ: Developed inflorescence and rachis
(20cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI;
An unidirectional cDNA library generated from [developed
inflorescence and rachis (20cm-long)]. cDNA was prepared
from polyA+ mRNA using Superscript Plasmid System Kit
(Invitrogen). The double-strand cDNAs were fractionated
in a sepharose CL-2B 40cm-columns and fragments sizing
between 0.8 and 1.5 Kb were directionally cloned into the
vector. Details of each source of RNA and library
construction can be obtained at
<http://sucest.lad.ic.unicamp.br/public>"

ORIGIN

Query Match 28.8%; Score 150.4; DB 14; Length 657;
Best Local Similarity 62.4%; Pred. No. 6.4e-12;
Matches 282; Conservative 0; Mismatches 146; Indels 24; Gaps 2;

```
QY 93 CACCACCGTTACGCCCACTCTCTGACGACCCCTTCTCCCTTCCACCGCGCTCTCC 152
DB 71 CTCACGCGGTGACGACGACCCCTTCTTCCCATCTTCCCTTCCACGACGACGAC 130
QY 153 TCCTCTGCGCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 212
DB 131 TCGTCTGCGCCCTTACATCTCGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 190
QY 213 GACTCCCTCTGCTGCTGACCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 272
DB 191 GACGACCCCTTCTTACCCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 250
QY 273 AGTCCCTCTGACCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 311
DB 251 ACCTTGACCTGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 310
QY 312 GACCGCGTCTCCCACTGAGCTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 371
DB 311 GACTTGACCCCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 370
QY 372 AGCAGGAAGTGACCTATGTACCGAGTCCACCGCGCATGTCAAGTGACCAAGGAG 431
DB 371 CGTCGCAAGTACACTACGCTCCGAGGCGGACGAGGAGGAGTCAAGTGACCGCGGAG 430
QY 432 GACAAGCTCGCGCGCGGAGAGGCTCTCAAGTGAGGCTGAGCTCTCTCTCTCTCT 491
DB 431 GACAGGCGC--CGCGCGGACCGGAGCTCAAGTGAGGAGCGGAGCTCAAGAGCGCTTAC 487
QY 492 GACGATGATTCGACCGCAAGTGAGAGTGGGA 523
DB 488 CACGACGCGCTTCGACCGCAAGTGAGAGTGGGA 519
```

RESULT 15

CA220794 667 bp mRNA linear EST 25-SEP-2003
LOCUS SCBGFL4050G08.g FL4 Saccharum officinarum cDNA clone SCBGFL4050G08
DEFINITION 5', mRNA sequence.
ACCESSION CA220794
VERSION CA220794.1 GI:35274458
KEYWORDS EST.
SOURCE Saccharum officinarum
ORGANISM Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum.
1 (bases 1 to 667)
Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.

REFERENCE
AUTHORS

Job time : 2733 secs

TITLE The libraries that made SUCEST
JOURNAL Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT Contact: Arruda P
Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parrruda@unicamp.br

Clone distribution: clone distribution information can be found
through the Brazilian Clone Collection Center (BCCC) at
http://www.bcccenter.ficav.unesp.br
Plate: 050 row: G column: 08
Seq primer: T7 Promoter Primer.

FEATURES
source 1. .667
Location/Qualifiers
/organism="Saccharum officinarum"
/mol_type="mRNA"
/db_xref="taxon:4547"
/clone="SCBGF14050G08"
/lab_host="DH10B"
/clone_lib="Fl4"

/note="Organ: Developed inflorescence and rachis
(20cm-long); Vector: pSport1; Site 1: SalI; Site 2: NotI;
An unidirectional cDNA library generated from [Developed
inflorescence and rachis (20cm-long)]. cDNA was prepared
from polyA+ mRNA using Superscript Plasmid System Kit
(Invitrogen). The double-strand cDNAs were fractionated
in a sepharose CL-2B 40cm-columns and fragments sizing
between 0.8 and 1.5 Kb were directionally cloned into the
vector. Details of each source of RNA and library
construction can be obtained at
http://sucest.lad.ic.unicamp.br/public"

ORIGIN

Query Match 28.8%; Score 150.4; DB 14; Length 667;
Best Local Similarity 62.4%; Pred. No. 6.4e-12;
Matches 282; Conservative 0; Mismatches 146; Indels 24; Gaps 2;

QY	93	CACCAACCGTTACGCCCACTCTCGACGACCCCTTCTCCCTTCCACCGCGCTCTCC	152
Db	82	CTCCACGCGCGTCGACGACGACCCCTTCTTCCCATCTTCCCTTCCCTACCAACACACC	141
QY	153	TCCTCTGCGCCCTCTCTCTCCACCGCGCGCTTCTCTCCACCTGCGCATCTTGGCCCTC	212
Db	142	TCGTCTGCGCCCTTACATCTCCGCTCTCTCTCTCCGCGCACCAACCGCTTCTCTCGAC	201
QY	213	GACTCCCCCTTCTGCTGCTGACCCCTTTCATCTCCACCCCTTCTCTCCCAAGCGGACACT	272
Db	202	GACCAACCCCTTCTACCCCACTCTCTGCGCGCTTGGCTTCACTCTCCCTCTCCCGATCGAC	261
QY	273	AGCTCCCTCTCGACCCCTTCTCTCTCCACACCTCAC-----C	311
Db	262	ACCTTCGACCTCGACTCTCTCTGCGCGCGCGCGCGCGCGCGCGCGGTGCTCGGCTTC	321
QY	312	GACCGCGCTTCTCCCACTCGAGCTCGCCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG	371
Db	322	GACTTCGACCCCTTCTCTCTCGACGCGCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCG	381
QY	372	AGCAGGAAGTGCACCTATGTACCGAGTCCACCGCGCGCGCGCATGTCAAGTGAACCAAGAG	431
Db	382	CGTCGCAAGTACACCTACGCTCCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG	441
QY	432	GACAAGCTTCGCGCGCGCGCGAGGGTGTCTCAAGTGGAGGCTGAGCTGACTTCCCTTAC	491
Db	442	GACAGGCG	498
QY	492	GACGATGATTCGACCGCGCAAGTGAAGTGGA	523
Db	499	CACGACGGCTTCGACCGCGCAAGTGAAGTGGA	530